Spring 2002

Robust Face Representation and Recognition Under Low Resolution and Difficult Lighting Conditions

Mohammad Moinul Islam
Old Dominion University

ROBUST FACE REPRESENTATION AND RECOGNITION UNDER LOW RESOLUTION AND DIFFICULT LIGHTING CONDITIONS

by

Mohammad Moinul Islam
B.Sc. March 2002, Bangladesh University of Engineering and Technology
M.Sc. December 2007, University of South Alabama

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL AND COMPUTER ENGINEERING

OLD DOMINION UNIVERSITY
May 2012

Approved by:

Mohammad A. Karim (Director)

Vijayan K. Asari (Co-Director)

Jiang Li (Member)

Gene Hou (Member)

ABSTRACT

ROBUST FACE REPRESENTATION AND RECOGNITION UNDER LOW RESOLUTION AND DIFFICULT LIGHTING CONDITIONS

Mohammad Moinul Islam
Old Dominion University, 2012
Director: Dr. Mohammad A. Karim
Co-Director: Dr. Vijayan K. Asari

This dissertation focuses on different aspects of face image analysis for accurate face recognition under low resolution and poor lighting conditions. A novel resolution enhancement technique is proposed for enhancing a low resolution face image into a high resolution image for better visualization and improved feature extraction, especially in a video surveillance environment. This method performs kernel regression and component feature learning in local neighborhood of the face images. It uses directional Fourier phase feature component to adaptively learn the regression kernel based on local covariance to estimate the high resolution image. For each patch in the neighborhood, four directional variances are estimated to adapt the interpolated pixels. A Modified Local Binary Pattern (MLBP) methodology for feature extraction is proposed to obtain robust face recognition under varying lighting conditions. Original LBP operator compares pixels in a local neighborhood with the center pixel and converts the resultant binary string to 8-bit integer value. So, it is less effective under difficult lighting conditions where variation between pixels is negligible. The proposed MLBP uses a two stage encoding procedure which is more robust in detecting this variation in a local patch. A novel dimensionality reduction technique called Marginality Preserving Embedding (MPE) is also proposed for enhancing the face recognition accuracy. Unlike Principal

Component Analysis (PCA) and Linear Discriminant Analysis (LDA), which project data in a global sense, MPE seeks for a local structure in the manifold. This is similar to other subspace learning techniques but the difference with other manifold learning is that MPE preserves marginality in local reconstruction. Hence it provides better representation in low dimensional space and achieves lower error rates in face recognition. Two new concepts for robust face recognition are also presented in this dissertation. In the first approach, a neural network is used for training the system where input vectors are created by measuring distance from each input to its class mean. In the second approach, half-face symmetry is used, realizing the fact that the face images may contain various expressions such as open/close eye, open/close mouth etc., and classify the top half and bottom half separately and finally fuse the two results. By performing experiments on several standard face datasets, improved results were observed in all the new proposed methodologies. Research is progressing in developing a unified approach for the extraction of features suitable for accurate face recognition in a long range video sequence in complex environments.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Karim for supporting me and giving suggestions during the period of study. I am grateful to Dr. Asari for his support and motivation in doing research. I would like to thank Dr. Li and Dr. Hou for spending their valuable time and suggestions. I am grateful to Dr. Anwer for arranging the dissertation defense without which it would be difficult. Finally, I would like to thank all the members of the Department and Research Foundation for their co-operation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 - INTRODUCTION

1.1 Problem Statement

This dissertation addresses four main topics: low resolution face enhancement, face representation under difficult lighting conditions, face recognition by reducing dimensionality using manifold learning and face recognition under varying expressions. Resolution enhancement refers to the process of obtaining a high resolution image from an image or a sequence of low resolution images. It has recently become a growing area of research in digital imaging and computer vision applications. In many applications, low cost imaging sensors are often used due to cost effectiveness which results in low quality, low resolution images. Increasing the resolution of the sensor is also not feasible due to the fact that shot noise increases during acquisition as the pixel size becomes smaller [1]. Therefore, super-resolution techniques can be used as an alternative to increase resolution.

Face recognition, although not a new area of research, still attracting a lot of researchers for its wide range of applications and the challenges that occur in real world environments. In practice, face images are obtained from different sources (ex. Facebook, Flickr, etc.) and at different times causing pose, appearance and illumination variations. So, a key challenge in face recognition problems is to find an efficient descriptor which is robust enough to handle these unconstrained conditions.

Another issue in computer vision and machine learning is that we often need to deal with very high dimensional data but the intrinsic structure of the data may lie in a low dimensional space. Learning such high dimensional data is computationally expensive

and not suitable for all practical applications. Moreover, it is also desirable to reduce the dimension for visualization. But this low dimensional data must preserve the underlying structure of high dimensional data in order to be of use. This leads researchers to develop methods of dimensionality reduction that can extract manifold structure of data on which data may reside. Another issue in face recognition is expression variations. It is a challenging work because human faces vary in pose and expression. We deal these problems with neural network based training systems.

1.2 Literature Review

The methods of super-resolution can broadly be classified into two categories: i) multiple frame super-resolution and ii) single frame super- resolution. Multiple frame super-resolution (ex., [2-6]) uses a set of low resolution scenes that are related by sub-pixel displacements. For static scenes, these sub-pixel displacements are global due to relative positions of the cameras and camera motion while for dynamic scenes they are local due to the object motion. These low resolution images can be used to reconstruct a high resolution image by employing proper motion estimation. But practically, this method is useful to small increase in resolution [7]. Super-resolution restoration in frequency domain has been proposed by many researchers [3-4]. Tsay et al. [6] reconstructed a high resolution image from several low resolution images using spatial aliasing effect.

Single image super-resolution methods are mainly based on interpolation and machine learning techniques. The problem of interpolation technique is that it smoothes edges and causes blurring problems. In order to avoid edge smoothness, a large number of edge based methods have been proposed in the literature [8-11]. Early edge preserving

methods depend on directional edges and interpolate the unknown pixels using neighborhood pixels. Li et al. proposed a new edge directed interpolation using covariance estimation [10]. The method uses geometric duality to establish the link between high resolution and low resolution covariance matrices. Another approach of edge guided interpolation [11] is to use pixel values in the two diagonal directions of a neighborhood and then fuse them to obtain minimum mean square error. All these methods suffer from aliasing effect for resolution factor above two. Recently, use of image prior information [12-14] (e.g., edge, corners and ridges) has become popular in super-resolution and restoration problem. Sun, et al. [12] proposed a different way of using edge information where parametric prior for image gradient is combined with a constraint to solve the optimization problem. Dai, et al. [13] used soft edge smoothness prior for color image super-resolution using alpha matting technique to utilize all color information from different channels by their alpha channel descriptors.

Many face representation techniques have been proposed so far including the Gabor feature Liu et al.[29], Lei *et. al* [31], principal component analysis (PCA), Turk et al., modified PCA [32], 2D PCA Yang *et. al.*, [33], Fisher's linear discriminant analysis (FLDA) Belhumeur, et. al., [34], independent component analysis (ICA) Liu et al., [35] Comon [36] etc. All these methods have been widely investigated and found to perform well under controlled settings. Recently, local texture descriptor using LBP has been shown to be effective in face recognition Tan et al., [37] Ahonen, *et. al.*, [38]. It has been used in combination with other descriptors such as Gabor, histograms etc. Zhang, *et. al.*, [39] Xie, *et. al.* [40] in order to improve recognition accuracy but a little attention is given to the improvement of original LBP operator. Zhao and Pietikäinen proposed

volume local binary patterns (VLBP) for dynamic texture recognition which extracts textures in spatiotemporal domain by applying LBP in three orthogonal directions Zhaoe et al., [41]. Another extension Lei, *et. al.*, [30] was conducted on Gabor face volume to explore the neighboring relationship in spatial, frequency and orientation domains. Wolf, *et al.* [54] proposed three-patch and four-patch LBP codes where the center pixel in a 3×3 neighborhood is encoded using eight (for three-patch and 16 for four-patch) additional 3×3 patch and the distance between two patches is thresholded to estimate the corresponding bit value. Tan et al. [37] quantized LBP to three levels namely local ternary patterns (LTP) in order to reduce noise effects in near uniform regions. All these methods perform well under small perturbation of lighting conditions.

Many techniques (both supervised and unsupervised) for dimensionality reduction have been proposed over the last few decades [29-33]. All these methods have validated one thing in common-recognition rate can significantly be improved at low dimensional subspace. Two of the most primitive techniques for this purpose are PCA [29-31] and LDA [30-33]. PCA transforms original image space to orthogonal feature space in the sense of mean square error. LDA seeks for linear transformations that minimize within class covariance and maximize between class covariance matrices. Unlike PCA, LDA encodes discriminating features in a linearly separable space that are not necessarily orthogonal [34]. When number of training data are small, PCA can outperform LDA but if the class information is available, LDA can be used to find optimal subspace for optimal discrimination [35]. Recently various research [34, 36-38] on face images have shown that data may reside on a nonlinear submanifold. As a result manifold learning becomes popular for face recognition. Some popular nonlinear techniques include

Laplacian Eigenmap [39], Locally Linear Embedding (LLE) [36] and Isomap [40]. All these methods showed impressive results on artificial datasets and some real applications. But they are defined only on training data points and it is unclear how the map can be evaluated for new data points. Some linear manifold learning techniques have also been proposed such as Locality Preserving Projection (LPP) [38] and Augmented Relation Embedding (ARE) [43]. LPP uncovers manifold structure by preserving local structure of data while ARE learn manifold by using users feedback. Yan, *et al.* [42] proposed Marginal Fisher Analysis (MFA) for dimensionality reduction. It redefines LDA in a graph embedding framework by constructing two graphs - one for intraclass compactness and the other for interclass separability. Another graph embedding network called Maximum Margin Projection (MPP) [43] maximizes the margin between within-class graphs and between-class graphs. However, previous work on face representation requires preprocessing to adjust illumination effect. Dimensionality reduction using manifold learning techniques were also not investigated from local reconstruction and class level simultaneously. These are the motivation of this dissertation.

1.3 Scope of the Dissertation

The contributions of this dissertation are four-fold:

- A feature based covariance estimation for adaptive kernel regression is proposed. Component feature learning is introduced and shows how it improves performance of the algorithm overtaking the feature as a whole.
- An improvement of Local Binary Pattern (LBP) is presented for robust face representation under varying lighting conditions. An original LBP operator

compares pixels in a local neighborhood with the center pixel and converts the resultant binary string to an 8-bit integer value. So, it is less effective under difficult lighting conditions where variation between pixels is negligible. The proposed MLBP in this dissertation uses two stage encoding procedure which is more robust in detecting this variation in a local patch.

- A novel dimensionality reduction technique called Marginality Preserving Embedding (MPE) is proposed. Unlike Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) which project data in a global sense, MPE seeks for local structure in the manifold. This is similar to other subspace learning techniques but the difference with them is that MPE preserves marginality in local reconstruction.
- A neural network based half-face similarity matrix is proposed for face recognition under varying expressions. Instead of taking the image vector as the network input, a vector is created by considering distance of each pattern from all class mean.

CHAPTER 2 - LOW RESOLUTION FACE ENHANCEMENT

2.1 Introduction

This chapter focuses on resolution enhancement of low resolution face image. In some applications (ex. video surveillance) face images are of low resolution. It is therefore necessary to increase the resolution without blurring the face. Toward this end, a neighborhood dependent component feature learning (NDCFL) technique is proposed to increase the resolution from a single image. Given a low resolution input, the method uses directional Fourier phase feature component to adaptively learn the regression kernel based on local covariance to estimate the high resolution image. Although this formulation resembles other regression and covariance based methods, this method uses image features to learn the local covariance from geometric similarity between low resolution image and its high resolution counterpart. For each patch in the neighborhood, four directional variances are estimated to adapt the interpolated pixels. Experimental results show that the proposed algorithm performs better than other state of the art techniques, especially at higher resolution scales.

2.2 Kernel Regression in Image Super-resolution

The problem of single image super-resolution can be stated as estimating a high resolution image I_h from a given low resolution image I_l. Mathematically low resolution images can be expressed as the decimated output of its high resolution counterpart using the following expression:

$$I_l\left[\frac{n_1}{r},\frac{n_2}{r}\right]=I_h[n_1,n_2] \qquad n_1,n_2=0,r,2r,... \tag{2.1}$$

where r is the resolution enhancement factor. The interpolated image can be obtained by convolving low resolution image with a discrete kernel, $K(n_1,n_2)$ which depends on the local characteristics of the image. This is a linear filtering process and the output image is found by combining the same weighted sample [21].

2.11 Kernel Regression

Kernel regression is a nonparametric method which depends on data values to find the structure of the model. It is also known as a weighted least squares solution and provides more freedom to interpolate using higher degree of polynomials. For two dimensional cases, the regression model can be defined as:

$$y_i=f(\mathbf{x}_i)+\varepsilon_i, \text{ for } i=1,2,...,P \text{ and } \mathbf{x}_i=[x_{1i},\ x_{2i}]^T \tag{2.2}$$

where $\{(\mathbf{x}_i), i=1,2,...,P\}$ are the sample locations, $\{y_i, i=1,2,...,P\}$ are the observations of the response variable $\mathbf{x}$, f is a regression function, $\{\varepsilon_i, i=1,2,...,P\}$ are independent identically distributed (i.i.d.) random errors, and P is the number of samples in the neighborhood. A generalized technique to estimate the unknown pixel values is to solve the following minimization problem [22].

$$\min_{\{\beta_n\}_{n=0}^{N}} \sum_{i=1}^{P}\left[y_i-\beta_0-\beta_1^T(\mathbf{x}_i-\mathbf{x})-\beta_2\,\mathrm{vech}\{(\mathbf{x}_i-\mathbf{x})(\mathbf{x}_i-\mathbf{x})^T\}-...\right]^2 K_{\mathbf{H}}(\mathbf{x}_i-\mathbf{x}) \tag{2.3}$$

with

$$K_{\mathbf{H}}(\mathbf{t})=\frac{1}{\det(\mathbf{H})}K\left(\mathbf{H}^{-1}\mathbf{t}\right) \tag{2.4}$$

where $K(.)$ is the kernel function which penalizes the samples based on their distance from the center of the kernel, $\beta_0, \beta_1, ... \beta_N$ are the unknown regression coefficients and $\mathbf{H}$ is a smoothing matrix of size 2×2. A simplified and computationally efficient model of smoothing matrix is defined as

$$\mathbf{H} = h\mu \mathbf{I} \tag{2.5}$$

where h is the global smoothing parameter and μ is a scalar constant (set to 1). A more detailed analysis about kernel regression can be found in [22-23]. To make this paper self-explanatory, the results are provided below. We can express (3) as a weighted least square optimization problem

$$\begin{aligned} \hat{b} &= \arg\min_{b} \left\| \mathbf{y} - \mathbf{X}_x \mathbf{b} \right\|^2_{\mathbf{W}_x} \\ &= \arg\min_{b} (\mathbf{y} - \mathbf{X}_x \mathbf{b})^T \mathbf{W}_x (\mathbf{y} - \mathbf{X}_x \mathbf{b}) \end{aligned} \tag{2.6}$$

where

$$\mathbf{y} = \left[y_1, y_2, ..., y_p \right] \tag{2.7}$$

$$\mathbf{b} = \left[\beta_0, \boldsymbol{\beta}_1^T, ..., \boldsymbol{\beta}_N^T \right]^T \tag{2.8}$$

$$\mathbf{W}_{\mathbf{x}} = \mathrm{diag}\left[K_{\mathbf{H}}(\mathbf{x}_1 - \mathbf{x}), K_{\mathbf{H}}(\mathbf{x}_2 - \mathbf{x}), .., K_{\mathbf{H}}(\mathbf{x}_p - \mathbf{x}) \right] \tag{2.9}$$

$$\mathbf{X}_{\mathbf{x}} = \begin{bmatrix} 1 & (\mathbf{x}_1 - \mathbf{x})^T & \mathrm{vech}^T\left\{ (\mathbf{x}_1 - \mathbf{x})(\mathbf{x}_1 - \mathbf{x})^T \right\} & \cdots \\ 1 & (\mathbf{x}_2 - \mathbf{x})^T & \mathrm{vech}^T\left\{ (\mathbf{x}_2 - \mathbf{x})(\mathbf{x}_2 - \mathbf{x})^T \right\} & \\ \vdots & \vdots & \vdots & \vdots \\ 1 & (\mathbf{x}_p - \mathbf{x})^T & \mathrm{vech}^T\left\{ (\mathbf{x}_p - \mathbf{x})(\mathbf{x}_p - \mathbf{x})^T \right\} & \cdots \end{bmatrix} \tag{2.10}$$

The least squares estimation of (6) is given by

$$\hat{f}(x) = \mathbf{e}_1^T \left(\mathbf{X}_\mathbf{x}^T \mathbf{W}_\mathbf{x} \mathbf{X}_\mathbf{x} \right)^{-1} \mathbf{X}_\mathbf{x}^T \mathbf{W}_\mathbf{x} \mathbf{y} \tag{2.11}$$

Here, $\mathbf{e}_1$ is a column vector with the first element equal to one and the rest equal to zero. The choice of kernel function is important because it affects the local estimation. Gaussian kernels have the property of decaying value with increasing distance and hence makes the estimation truly local. Takeda, *et al.* [22] proposed a data adapted steering kernel which is defined as follows:

$$K_{\mathbf{H}_i}(\mathbf{x}_i - \mathbf{x}) = \frac{\sqrt{\det(\mathbf{C}_i)}}{2\pi h^2 \mu_i^2} \exp\left\{ -\frac{(\mathbf{x}_i - \mathbf{x})^T \mathbf{C}_i (\mathbf{x}_i - \mathbf{x})}{2h^2 \mu_i^2} \right\} \tag{2.12}$$

The advantage of this kernel is that in the exponential term it uses Mahalanobis distance which is a generalization of Euclidean distance but it can adaptively point or steer along the data structure [24]. Steering kernel was is used throughout the dissertatin. The data dependent parameter in steering kernel is covariance matrix, $\mathbf{C}_i$ which is estimated locally by decomposing image features and geometry of the image. The following section is devoted to the estimation of covariance matrix.

2.3 Feature Based Covariance Estimation

Image super-resolution problem has been studied for years but a little attention is made to feature based super-resolution. In most of the works related to super-resolution, emphasis is given to preserve edges or high frequency details. So, edge or gradient information is used to predict high resolution image. But in case of texture features, it may not always a

good idea to analyze in spatial domain where image gradient hardly convey any useful information. Here, a feature based covariance estimation is proposed in space-frequency domain which use local Fourier transform and is more robust in preserving details with changing scale factor. Single image super-resolution is generally an ill-posed problem because of insufficient data and unknown blurring operator [19]. A common trend is therefore to find correlation between low resolution patch and its high resolution counterpart which determines the reconstruction error. It has been observed that component features improve face detection and recognition rate than taking the features as a whole [25]. This inspires the proposal of a component feature decomposition technique in this literature. Although any image decomposition techniques are equally applicable, that is restricted to Fourier phase features only. Fourier feature has been widely used in many image processing applications because of its simplicity and ease of operation. The Fourier transform of image, $I_l(x_1, x_2)$ is defined as:

$$F(u,v) = \frac{1}{YZ}\sum_{x_1=0}^{Y-1}\sum_{x_2=0}^{Z-1} I_l(x_1,x_2)\exp\left[-j2\pi\left(\frac{ux_1}{Y}+\frac{vx_2}{Z}\right)\right] \tag{2.13}$$

where $0 \le u \le Y-1$ and $0 \le v \le Z-1$. Fourier transform is a complex representation of a spatial image. Real and imaginary components of this transformation can be expressed in terms of magnitude and phase as:

$$\mathrm{Re}(u,v) = |F(u,v)| \times \cos[\varphi(u,v)] \tag{2.14}$$

$$\mathrm{Im}(u,v) = |F(u,v)| \times \sin[\varphi(u,v)] \tag{2.15}$$

where $|F(u,v)|$ and $\varphi(u,v)$ are magnitude and phase respectively. Since magnitude spectrum does not contain any significant feature, only consider phase spectrum is considered. The phase components corresponding to real and imaginary parts of transformation are as follows:

$$P_{\varphi} = \cos[\varphi(u,v)] \tag{2.16}$$

$$Q_{\varphi} = \sin[\varphi(u,v)] \tag{2.17}$$

The covariance matrix is estimated separately for P_{φ} ($\mathbf{C}_{1i}$) and Q_{φ} ($\mathbf{C}_{2i}$)components. So, the local covariance matrix is calculated as

$$\mathbf{C}_i = \mathbf{C}_{1i} + \mathbf{C}_{2i} \tag{2.18}$$

2.3.1 Covariance Matrix Using Directional Variance

This section demonstrates how the phase feature is incorporated with local geometry and statistics to compute covariance matrix. It is assumed that for 3X magnification a l resolution image, I_l is obtained directly by down-sampling a high resolution image I_h which isrelated by $I_l(x_1,x_2)=I_h(4x_1-3,4x_2-3)$, $1\le x_1 \le Y$, $1\le x_2 \le Z$. Fig. 2.1(a) shows a schematic block diagram of the algorithm and Fig. 2.1(b) refers to a 9×9 local neighborhood patch. This 9×9 patch are taken from input LR image and considered an HR patch and the black dots in it (3×3 patch) are considered a LR patch. So, the super-resolution problem is to estimate unknown pixels using kernel regression technique as discussed in the previous section. In the kernel function the only data dependent parameter is its covariance matrix. The quality of the final output depends on how the covariance matrix is devised. Since the human visual system is more sensitive to edges, it

is important to estimate the covariance along these directions. Although the idea is not new, Li and Orchard [10] shows "new edge directed interpolation (NEDI)" technique using "geometric duality" which refers to the correspondence between low resolution covariance and high resolution covariance along the same direction. So, they estimate high resolution covariance from its low resolution counterpart. The problem here is that the method uses linear interpolation of corner edge pixels, so at higher resolution like 4x, artifacts become prominent near the edge areas.

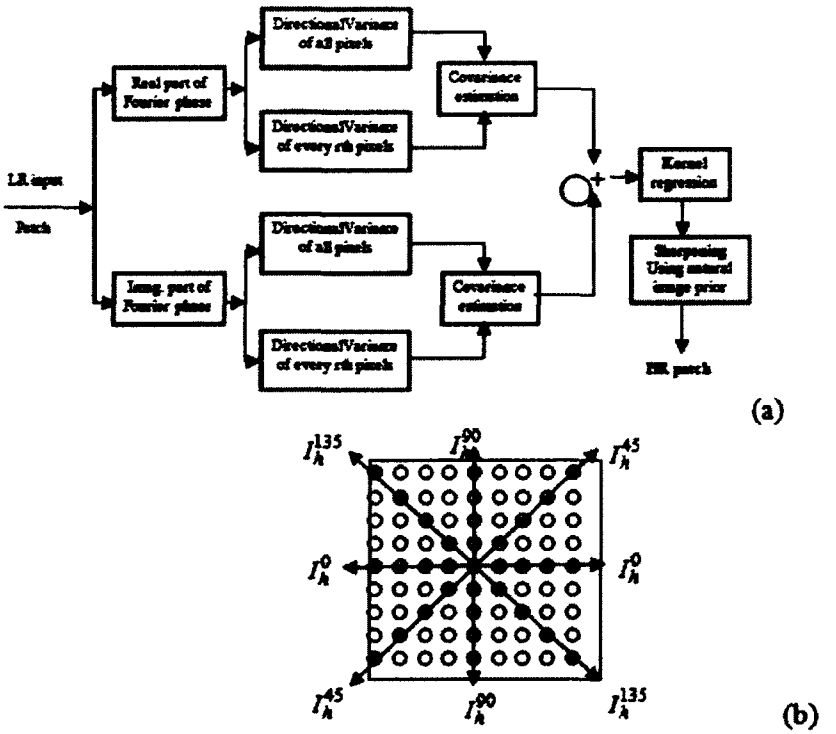


Figure 2.1 (a) Block diagram representation of the proposed super-resolution algorithm (b) Local neighborhood of the center dark pixel. For high resolution patch all 9 pixels are taken in each direction while for low resolution patch (of order 3) only dark pixels are used.

In this dissertation, covariance is estimated in the frequency domain. In the frequency domain, spatial frequency of image corresponds to the rate of its intensity variation. Higher frequencies are found near the region of higher fluctuation of intensity whereas lower frequencies are concentrated near the lower intensity variation. So, if we take Fourier transform of a high resolution image, its spatial frequency reflects the variation of intensity throughout the image. Now, let us take a second image which is obtained by downsampling the first image and then upsampling to the same size by pixel replication. Fourier transform of the second image will also give spatial frequency which reflects the intensity variation of low resolution image. So, we can establish correlation between these two in their transformed domain. But taking spatial frequency from the second image will suffer from aliasing effect because only every r^{th} (resolution factor) pixel contains new information, so in an analogous manner every r^{th} pixel was taken from the first image.

Another issue in conventional covariance estimation is using image intensity or gradient information. To exemplify, consider a high resolution image and its reconstruction from a downsampled image using pixel replication. The reconstructed image suffers from an aliasing effect because of pixel replication. A measure of goodness of a super-resolution image is thus determined by the variability or distribution of intensity values in a neighborhood of a given pixel. With that intuition in mind, directional variance mapping is proposed between a low resolution image and its high resolution counterpart. Lei and Xiaolin [11] use directional interpolation and fuse them to estimate high resolution pixel value. The difference with their method is that here directional variance is used to estimate covariance between low resolution and high

resolution patch which in turn steer the kernel. So, it is dependent on image structure. The given low resolution image is considered a high resolution truth mask. For each pixel location of the LR input image, a 9×9 overlapping patch is created and variances of its feature values along horizontal, vertical and two diagonal directions are measured as:

$$\mathrm{Var}\left(I_h^o\right)=\frac{1}{n}\sum_{k=1}^{n}\left(I_h^0(k)-\mu_h^0\right)^2 \qquad \mu_h^0=E\left[I_h^0\right] \tag{2.19}$$

$$\mathrm{Var}\left(I_h^{45}\right)=\frac{1}{n}\sum_{k=1}^{n}\left(I_h^{45}(k)-\mu_h^{45}\right)^2 \qquad \mu_h^{45}=E\left[I_h^{45}\right] \tag{2.20}$$

$$\mathrm{Var}\left(I_h^{90}\right)=\frac{1}{n}\sum_{k=1}^{n}\left(I_h^{90}(k)-\mu_h^{90}\right)^2 \qquad \mu_h^{90}=E\left[I_h^{90}\right] \tag{2.21}$$

$$\mathrm{Var}\left(I_h^{135}\right)=\frac{1}{n}\sum_{k=1}^{n}\left(I_h^{135}(k)-\mu_h^{135}\right)^2 \qquad \mu_h^{135}=E\left[I_h^{135}\right] \tag{2.22}$$

where n is the number of pixels in each direction (for I_h, $n=9$ and for I_l, $n=3$) and μ_h is the expected value in corresponding direction. Similarly, for low resolution patch $\mathrm{Var}\left(I_l^0\right)$, $\mathrm{Var}\left(I_l^{45}\right)$, $\mathrm{Var}\left(I_l^{90}\right)$ and $\mathrm{Var}\left(I_l^{135}\right)$ are estimated using black dots only as shown in Fig. 2.1(b). Now, the covariance matrix C_{fi} ($f=$ 1, 2 represents real and imaginary part of phase features respectively) and can be estimated as:

$$C_{fi}=\begin{bmatrix}\mathrm{Var}\left(v_h\right) & c\sqrt{\mathrm{Var}\left(v_h\right)\cdot\mathrm{Var}\left(v_l\right)}\\ c\sqrt{\mathrm{Var}\left(v_h\right)\cdot\mathrm{Var}\left(v_l\right)} & \mathrm{Var}\left(v_l\right)\end{bmatrix} \tag{2.23}$$

where

$$v_h=\left[\mathrm{Var}\left(I_h^0\right)\ \mathrm{Var}\left(I_h^{45}\right)\ \mathrm{Var}\left(I_h^{90}\right)\ \mathrm{Var}\left(I_h^{135}\right)\right]^T \tag{2.24}$$

$$v_l = \left[\mathrm{Var}\left(I_l^0\right), \mathrm{Var}\left(I_l^{45}\right), \mathrm{Var}\left(I_l^{90}\right), \mathrm{Var}\left(I_l^{135}\right)\right]^T \tag{2.25}$$

and c is the normalized correlation co-efficient defined as:

$$c = \frac{E\left[(v_h - \mu_h)\cdot(v_l - \mu_l)\right]}{\sqrt{E\left[(v_h - \mu_h)^2\right]\cdot E\left[(v_l - \mu_l)^2\right]}} \tag{2.26}$$

So, the local covariance matrix $\mathbf{C}_{fl}$ is of size 2×2. Covariances are estimated at all pixel locations and the final covariance matrix is a four dimensional matrix of size 2×2×Y×Z, where Y×Z is the size of the input image. The techniques discussed here remove artifacts and blur effects near edge areas which are present in most of the interpolation techniques for resolution factor of 3 and above.

2.3.2 Algorithm

1. Input: a low resolution image and the resolution enhancement factor r.
2. For each patch of size 9×9,
    a) Calculate real components of Fourier phase angle and normalize them.
    b) Calculate four variances (4×1 vector) taking all 9 pixels along horizontal, vertical and two diagonal directions about the center pixel as shown by hatched pixels in Fig. 2.1(b).
    c) For the same patch, estimate four variances (4×1 vector) taking only the r^{th} pixels along the same (respective) directions.
    d) Calculate covariance of the two vectors. This gives covariance matrix at the center pixel.
3. Repeat step 2 for imaginary components of Fourier phase angle.

4. Add two covariance matrices found in step 2 and step 3. Substitute in Eq. (2.12) to calculate the kernel function.
5. Use Eq. (2.11) to estimate the unknown pixels.
6. Output: high resolution image of order r.

2.4 Simulation Results

In the experiments, the algorithm was mostly tested to magnify input images for resolution factor of 3 and 4. In estimating the covariance matrix, a 9×9 overlapping patch is always used, as shown in Fig. 1(b). For high resolution patch, variances are estimated by taking all 9 pixels in each direction but for low resolution patch of order 4, every 5^{th} pixel was taken. The two feature images (cosine and sine components) are then used to estimate patch based local covariance matrices. In kernel regression, the sum of the above covariance matrices is used. There are two parameter settings in the kernel regression problem - one is the kernel size k and the other is the smoothing parameter h. Using a large kernel size tends to blur the image and a high value of h smooth out the image. Fig. 2.4 shows plots of RMSE with respect to smoothing parameter, h for different size of kernel, k. From the plot, it is evident that minimum RMSE is observed for $k = 3$ and $h = 0.5$. The original images are first decimated and then reconstructed using the method used in the dissertation and some of the state of the art methods.

In Fig. 2.2, the algorithm is compared on the image of the girl with other popular methods including NEDI [10], fusion [11] and ICBI [27]. The result of NEDI is sharp in appearance but generates high frequency artifacts throughout the face. The fusion result is over-smooth and the ICBI technique has fewer artifacts but fails to recover small

details. On the other hand, the method used in this dissertation generates fine details all over the image without artifacts. Fig. 2.3 shows more results on the Lena image using neighbor embedding (NE) [17], ICBI [27] and proposed methods. Quantitative measures are also used to compare the performance of the algorithm with others. The widely used root mean square error (RMSE) is computed for all images1. But a lower RMSE does not necessarily correspond to a better image. So the mean square similarity (MSSIM) [28] indexes (with window size 13×13) is also measured for all images. Table 2.1 shows the RMSE comparison and Table 2.2 shows the MSSIM comparison of the proposed method with other state of the art methods for a resolution factor 4. From the tables it can be seen that the method used in this dissertation always performs better than other methods which verify the structural similarity with original images.

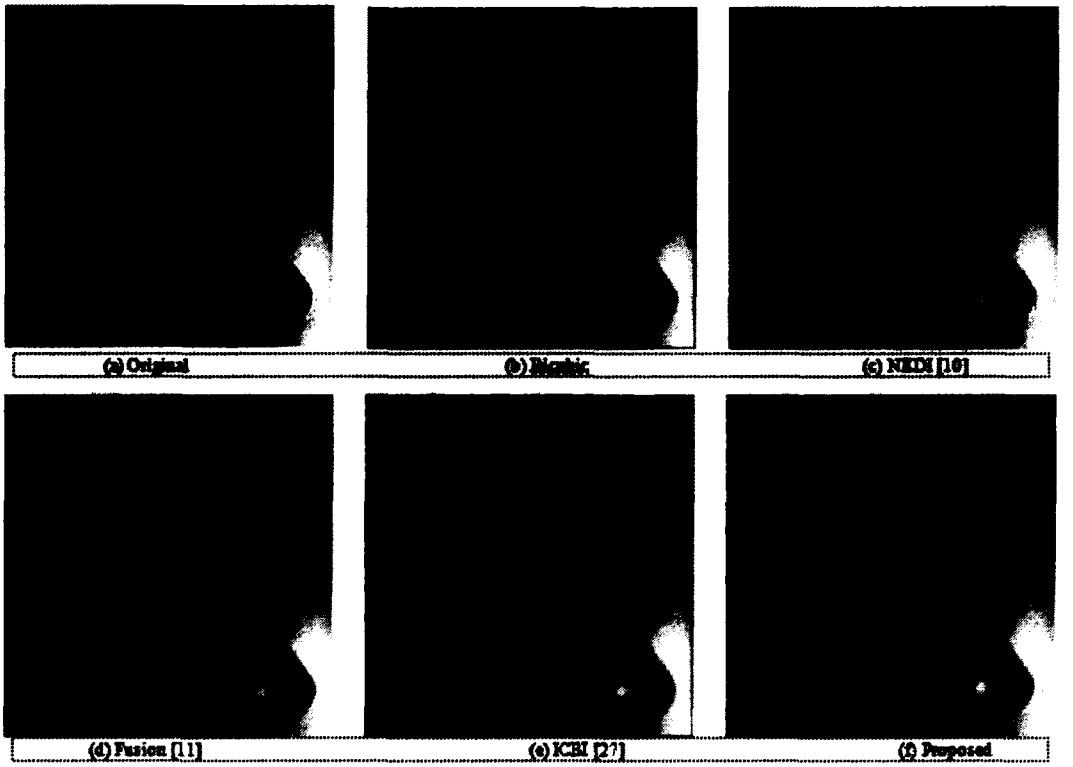


Figure 2.2: Comparison of girl image for 4x magnification with bicubic, NEDI [10], data fusion [11] and ICBI [27].

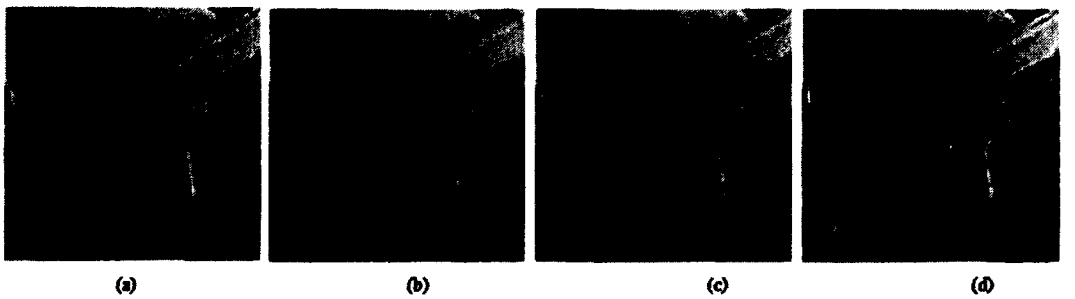


Figure 2.3: Comparison of girl image for 4x magnification with bicubic, NEDI [10], data fusion [11] and ICBI [27]. (a) Original image (b) Result of NE [17] (c) Result of ICBI [27] and (d) Proposed method.

Table 2.1 : RMSE comparison using correctness measures from the same ground truth image on 4X magnification.

	Bicubic	NEDI [10]	NE [17]	ICBI [27]	Proposed
Girl	12.1860	12.2083	10.2913	10.1881	**10.0904**
Lena	15.8843	18.2197	14.5722	13.9540	**13.8323**

Table 2.2: MSSIM comparison using correctness measures from the same ground truth image on 4X magnification.

	Bicubic	NEDI [10]	NE [17]	ICBI [27]	Proposed
Girl	0.5678	0.5593	0.5873	0.6276	**0.6486**
Lena	0.6850	0.5892	0.6907	0.7249	**0.7448**

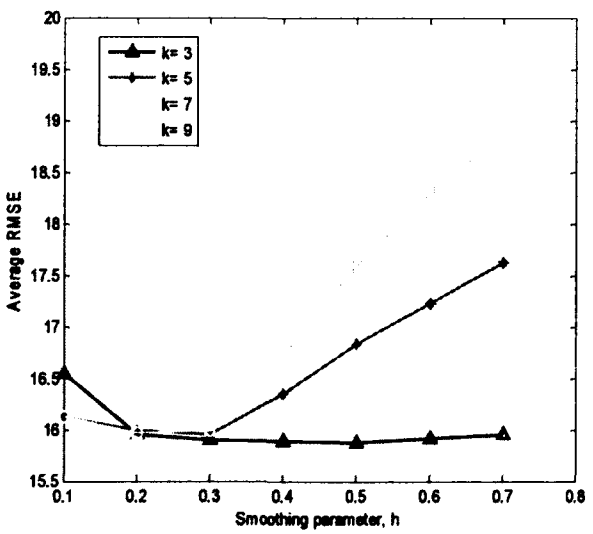


Figure 2.4: Plot of average RMSE with smoothing parameter for different kernel size.

2.5 Conclusions

In this chapter, a novel technique for single image super-resolution is presented. The proposed algorithm is a hybrid edge and feature based technique which uses image features in frequency domain. The covariance matrix is estimated component wise - one from learning real parts and the other from imaginary parts and then accumulating the results. We noticed that RMSE of the Lena image has improved from 14.0422 (for 4X without component learning) to 13.8323 (for 4X with component learning). Unlike other edge based methods, kernel regression technique was used to interpolate the unknown pixels and the kernel is learned from local features of the image. This enables noise reduction and artifacts while maintaining the sharpness of the image. Conventional super-resolution techniques use intensity or gradient information in estimating local covariance matrix. This chapter introduces the concept of directional variance. This tells us how the pixel values are distributed within the neighborhood of the image which is an important factor of image super-resolution. Experimental results show that proposed algorithm performs better than other resolution enhancement techniques.

CHAPTER 3 - FACE REPRESENTATION USING MODIFIED BINARY PATTERN

3.1 Introduction

FACE recognition, although not a new area of research, still attracts a lot of interest because of its wide range of applications and the challenges that occur in real world environments. In practice, face images are obtained from different sources (ex. Facebook, Flickr, etc.) and at different times causing pose, appearance and illumination variations. So, a key challenge in face recognition problems is to find an efficient descriptor which is robust to these unconstrained conditions. Many face representation techniques have been proposed so far including the Gabor feature by Liu et al.[29] and Lei et. al [31], principal component analysis (PCA) and modified PCA by Turk et al.[32], 2D PCA by Yang et. al., [33], Fisher's linear discriminant analysis (FLDA) by Belhumeur, et. al., [34], independent component analysis (ICA) by Liu et al., [35] and Comon [36] etc. All these methods have been widely investigated and found to perform well under controlled settings.

Recently, local texture descriptor using local binary pattern (LBP) has been shown to be effective in face recognition by Tan et al., [37] and Ahonen, et. al. [38]. It has been used in combination with other descriptors (such as Gabor, histogram, etc. [39-40]) in order to improve recognition accuracy but little attention is given to the improvement of original LBP operator. Zhao et al [41] proposed volume local binary patterns (VLBP) for dynamic texture recognition which extracts textures in spatiotemporal domain by applying LBP in three orthogonal directions. Another extension [30] is conducted on

Gabor face volume to explore the neighboring relationship in spatial, frequency and orientation domains. Wolf, *et al* [42] proposed three-patch and four-patch LBP codes where the center pixel in a neighborhood is encoded using 8 (for three-patch and 16 for four-patch) additional patches and the distance between two patches are thresholded to estimate the corresponding bit value. Tan et al [38] quantized LBP to three levels, namely local ternary patterns (LTP), in order to reduce noise effects in near uniform regions. All these methods perform well under small perturbation of lighting conditions.

This chapter addresses the illumination effects in feature extraction and a novel technique to improve conventional LBP coding is also proposed which can better handle lighting variations. Any small change or uniform texture pattern (which is the case in difficult lighting conditions) can easily be detected using the proposed technique and it is computationally efficient.

3.2 Review of Texture Coding Scheme

3.2.1 Local Binary Pattern (LBP)

The LBP operator, introduced by Ojala, *et al.* [43] is a powerful tool for texture description. It has been widely used in various recognition algorithms for its discriminative nature in texture classification. The LBP operator was originally defined for a 3×3 neighborhood and 8-bit binary pattern which gives $2^8 = 256$ possible texture units. It takes a neighborhood around each pixel and compares every pixel in the neighborhood with its centre pixel. The result of the comparison is then thresholded to give a binary number which is given a particular weight based on its position in the neighborhood. This gives an integer number of 8-bit LBP code around the center pixel

and is a local descriptor of that pixel. Figure 3.1 shows the LBP coding scheme and Fig. 3.2(a-c) shows images of a person with three different lighting conditions and their LBP and MLBP images are depicted in Fig. 3.2(d) and Fig. 3.2(e) respectively. The LBP coding of a 3×3 example patch with the center pixel as threshold is shown in Fig 3.1(f). Mathematically, LBP operator can be described as [44]

$$L(p_k - p_c) = \begin{cases} 1, & p_k \geq p_c \\ 0, & \text{otherwise} \end{cases} \tag{3.1}$$

$$\text{LBP} = \sum_{k=0}^{7} 2^k L(p_k - p_c) \tag{3.2}$$

Where p_k where $k = 0, 1, 2, ..., 7$) represents neighborhood pixels and p_c is the center pixel of that neighborhood. The LBP pattern of the pixel is calculated by assigning a binomial factor 2^k for each $L(p_k - p_c)$. The LBP operator was later extended to different sizes of neighborhood in order to deal with textures at different scales [55]. Another extension is the so-called uniform patterns (Zhao et al., 2007). A pattern is called uniform if it contains at most two bitwise transitions from '0' to '1' or vice versa in a circular fashion.

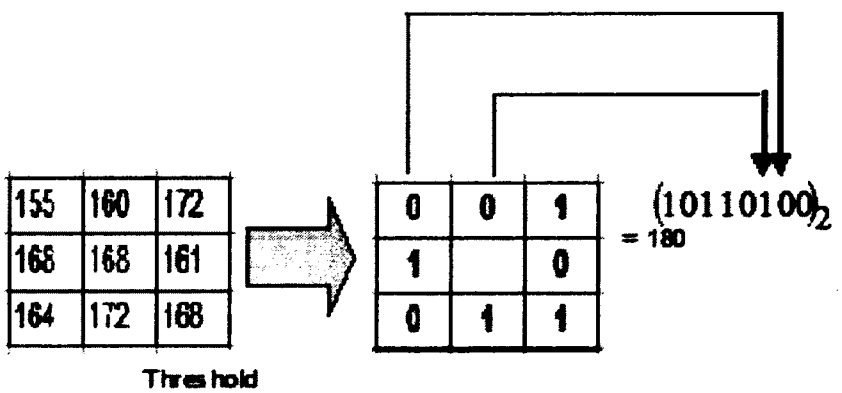


Figure 3.1: LBP encoding is shown for a 3×3 neighborhood with the center pixel as the threshold.

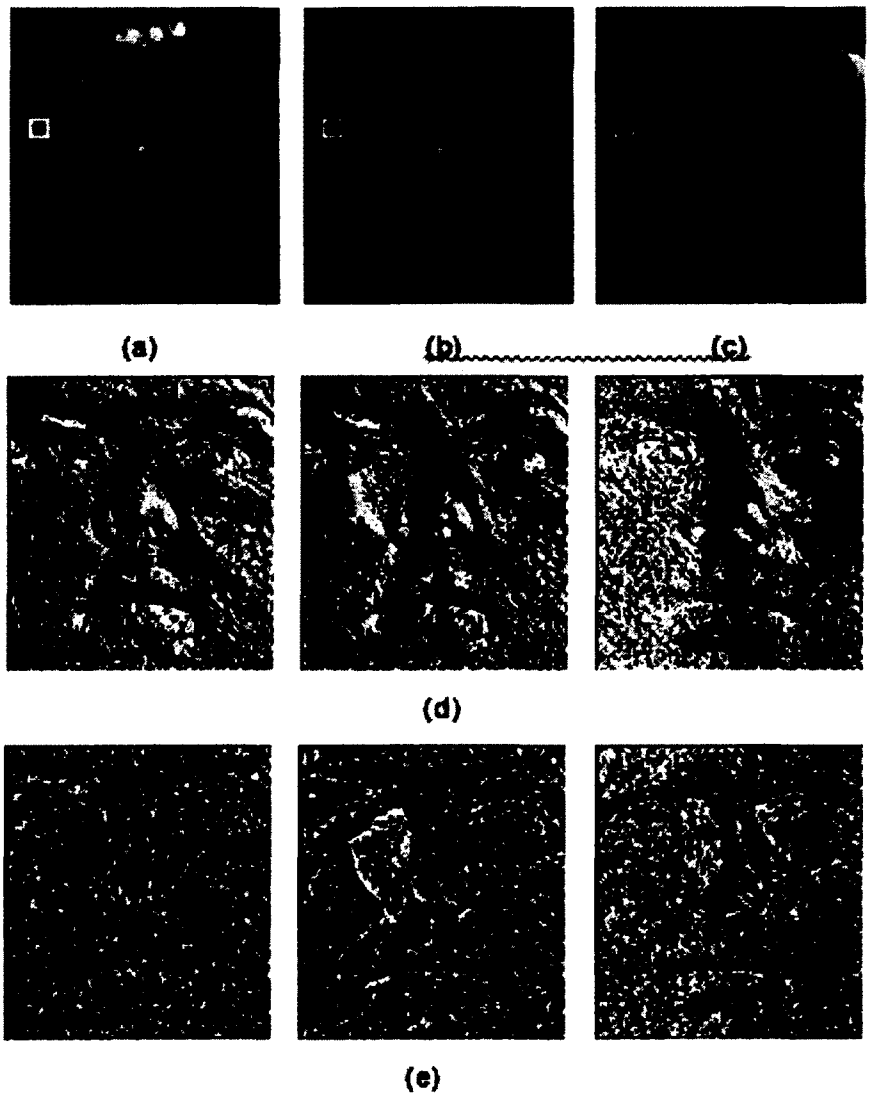


Figure 3.2: (a-c) An example image with three different illuminations and their corresponding (d) LBP image and (e) MLBP image.

3.2.2 Local Ternary Patterns (LTP)

LBP has highly discriminative features for texture classification and is resistant to lighting effects to some extent. However because they threshold at exactly the same value

of the center pixel, i_c they tend to be sensitive to noise. To reduce this effect, Local Ternary Patterns (LTP) is proposed [37] in which gray-levels in a zone of width $\pm t$ around i_c are quantized to zero, ones above the this are quantized to $+1$ and ones below it to -1. Mathematically, it can be expressed as

$$s'(u,i_c,t)=\begin{cases}1, & u \geq i_c + t \\ 0, & |u-i_c|<t \\ -1 & u \leq i_c - t\end{cases} \tag{3.3}$$

Here t is a user-specified threshold- so LTP codes are more resistant to noise but no longer strictly invariant to noise. LTP coding scheme is illustrated in Fig. 3.1. For simplicity, LTP coding scheme is split into positive and negative LBP codes as illustrated in Fig 3.2.

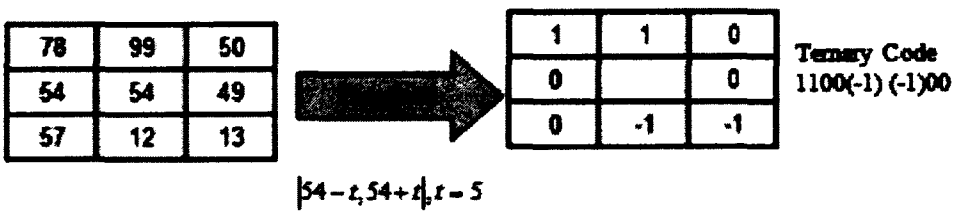


Figure 3.3: Illustration of the basic LTP operator.

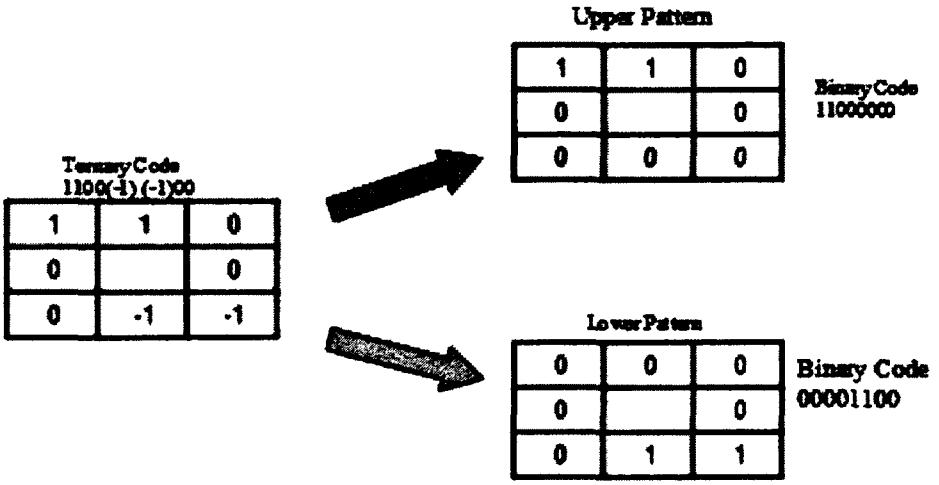


Figure 3.4 Splitting an LTP code into positive and negative LBP codes.

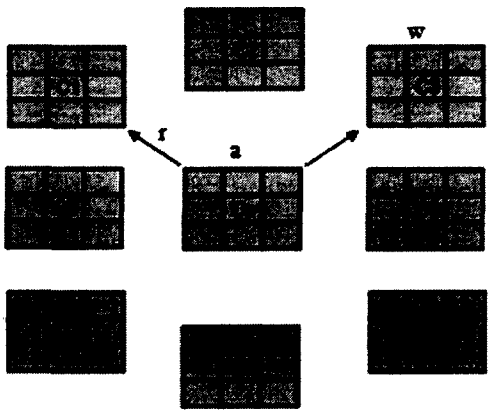
3.2.3 Three-Patch LBP (TPLBP)

In this section, three-patch LBP (TPLBP) codes are described that are produced by comparing the values of three patches to produce a single bit value in the code assigned to each pixel [54]. For each pixel in the image, a $w \times w$ patch centered on the pixel and S additional patches distributed uniformly in a ring of radius r around it as shown in Fig. 3.2 For a parameter α, pairs of patches separated α-patch apart are compared with the center pixel. The resulting code contains S bits per pixel computed as,

$$TPLBP_{r,S,w,\alpha}(p)=\sum_{i=1}^{S} f\left(d\left(C_i,C_p\right)-d\left(C_{i+\alpha \bmod S},C_p\right)\right)2^i \tag{3.4}$$

where C_i and $C_{i+\alpha \bmod S}$ are two patches along the ring and C_p is the central patch. The function $d(:,:)$ is any distance function between two patches and f is defined as:

$$f(x)=\begin{cases}1, & if\ x \geq \tau \\ 0, & if\ x < \tau\end{cases} \tag{3.5}$$



(a)

$$\begin{aligned}TPLBP_{r,S,2,\alpha}(p)=\ & f(d(C_0,C_p)-d(C_2,C_p))2^0+\\ & f(d(C_1,C_p)-d(C_3,C_p))2^1+\\ & f(d(C_2,C_p)-d(C_4,C_p))2^2+\\ & f(d(C_3,C_p)-d(C_5,C_p))2^3+\\ & f(d(C_4,C_p)-d(C_6,C_p))2^4+\\ & f(d(C_5,C_p)-d(C_7,C_p))2^5+\\ & f(d(C_6,C_p)-d(C_8,C_p))2^6+\\ & f(d(C_7,C_p)-d(C_1,C_p))2^7\end{aligned}$$

(b)

Figure 3.5 a) Three-patch LBP code with α = 2 and $S = 8$ b) TPLBP code with parameters S=2, α = 2 and w=3.

3.2 Modified Local Binary Pattern (MLBP)

The LBP operator has been successfully applied in many recognition applications for discriminative feature extraction. It has also proven its robustness in small change in lighting conditions. Since texture features are usually described by the relative change in pixel itensity with respct to its neighborhood and since the LBP operator compares and thresholds the neighborhood pixels at exactly the center pixel, it can extract the texture feature when there are significant variations in image intensity. But in case of difficult lighting conditions (extreme dark or bright), there are small variations in a local neighborhood; i.e, the variance of a local patch is negligibly small and pixel intensity repeats itself. So, there is a tendency that a '0' is encoded as '1' or vice versa when compared with an image with neutral light settings. As a result, the bit error rate increases and decoded value differs significantly.

In order to overcome the problem of LBP operator, a modified local binary pattern (MLBP) is proposed. This method uses two steps to encode the final pattern. First, a status bit is assigned to each pixel based on its local neighbors. Then, these status bits are used to encode the LBP of the center pixel. Figure 3.2 illustrates the proposed MLBP coding scheme. An input image is first converted to a binary status image. For each pixel a 3×3 (for MLBP3 and 5×5 for MLBP5) local neighborhood is selected and absolute intensity differences of all the pixels are calculated within the patch with respect to the center pixel. The sum of all deviations is denoted as total deviation (TD) for the centerpixel, P_c. Then all the pixels in the patch are comparedwith P_c and those pixels which are equal or above the center pixel are selected. Now, the deviation for these pixels

is estimated and denoted as the positive deviation (PD). The status bit of P_c is calculated as:

$$S_c = \begin{cases} 1, \text{ if } PD > \frac{1}{2}TD \\ 0, \text{ otherwise} \end{cases} \tag{3.6}$$

This gives a binary status image where each bit is estimated based on its local neighborhood. Now, this status image is taken as the input and a 3×3 neighborhood is taken to calculate MLBP code. The MLBP code of the center pixel is denoted using the status bits of all the neighboring pixels as shown in Figure 3.2. Thus the value of the center pixel is calculated by assigning a binomial factor as:

$$\text{MLBP} = \sum_{k=0}^{7} 2^k S_{k+1} \tag{3.7}$$

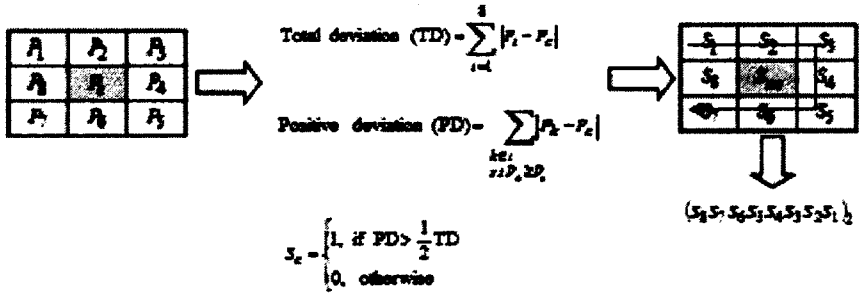


Figure 3.6: MLBP encoding; from left to right: a 3 × 3 neighborhood for calculating the status bit of the center pixel P_c, status bit calculation, MLBP code of center pixel is calculated from the status bits of its neighborhood.

In order to investigate robustness of the proposed method, histograms of three segments extracted from three images of the same person with different lighting are plotted. Three white rectangle regions are shown in Figures 3a, 3b and 3c to illustrate this and to obtain their LBP- and MLBP-generated histograms which are shown in Figure 3.4.

In Figure 3.3, image 'a' represents the most neutral lighting condition. Figure 3.4(a) shows original intensity plot of the images and they are at three different regions of the dynamic range of gray level. Figure 3.4(b) and Figure 3.4(c) plot histograms of their LBP and MLBP encoded images respectively. From the figures, it can be seen that LBP generated histograms are separated from each other significantly whereas their MLBP generated histograms resembles each other. This deviation is also measured quantitatively.

The L2-norm distances of image 'b' and image 'c' from image 'a' are calculated as 14.63 and 21.17 for LBP while it is more uniform in the case of MLBP, which are obtained as 12.65 and 12.73 respectively. Figures 3.3a to 3.3c show images of a person with three different lighting conditions and their LBP and MLBP images as depicted in Figure 3.3(d) and Figure 3.3(e), respectively.

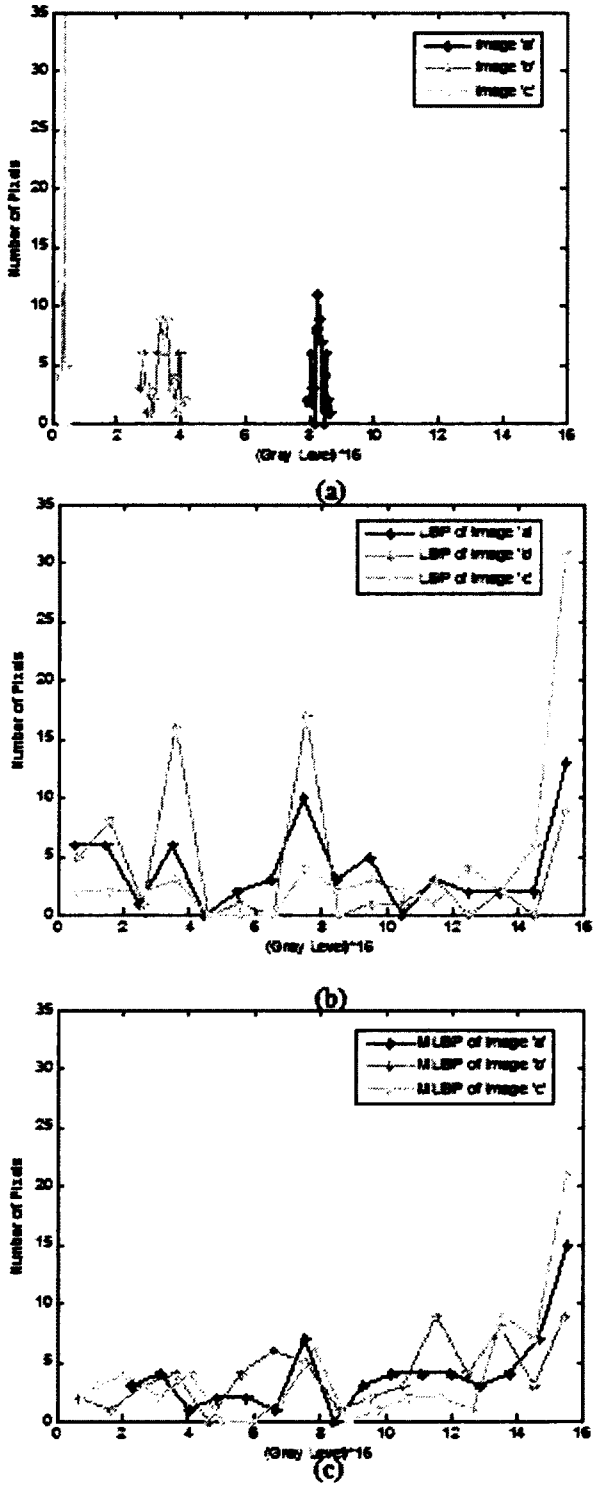


Figure 3.7: Histogram plot for three different illuminations (a) original intensity histograms (b) histograms of LBP images and (c) histograms of MLBP images.

3.2 Experimental Analysis

In this section the effectiveness of the proposed method is illustrated on the Extended Yale B database. The database contains 38 subjects under 64 illumination conditions. It has little variability in pose and expression but its extreme lighting variations make it a difficult problem in face recognition. From the database 2413 images of 38 individuals were selected. The images were cropped and resized to 36×30 pixels. The database was divided into two non-overlapping groups (group A and group B). Group A contains the subjects with odd numbered IDs (a total of 20 subjects) and Group B contains the remaining 18 subjects. For training purpose, two images were selected from each subject with the most neutral lighting conditions. The proposed was applied to the MLBP of each of the cropped face images and performed nearest neighbor (NN) classification in Euclidean space. The experimental results are listed in Table 3.1. It shows that LBP has higher error rate than MLBP and increasing the neighborhood size further improves recognition accuracy.

In this experiment the MLBP approach is compared with the LBP operator for determining recognition accuracy at various illumination conditions. For this purpose, the database is divided into six subsets according to their azimuth angles (5°, 20°, 35°, 50°, 85° and 120°). The performance of the LBP and MLBP algorithms are compared for each test image and the results are plotted in Figure 3.5. From the figure, it can be seen that at small lighting variations both the LBP and MLBP perform almost equally, but at strong lighting variations the performance of LBP degrades significantly.

At 85° and 120°, the proposed method outperforms the LBP by as much as 20%. In all these experiments MLBP5 shows better performance than both LBP and MLBP3. The

performance of LBP and MLBP is also compared at various dimensionalities. Figure 3.6 is a plot of recognition accuracy with dimensionality (row × column). This indicates that the performance of MLBP is much better than LBP at reduced dimensions.

Table 3.1 Comparison of the Performance on Error Rate Using Yale B Database.

Method	Error Rate
LBP-NN	0.14
MLBP3-NN	0.08
MLBP5-NN	0.05

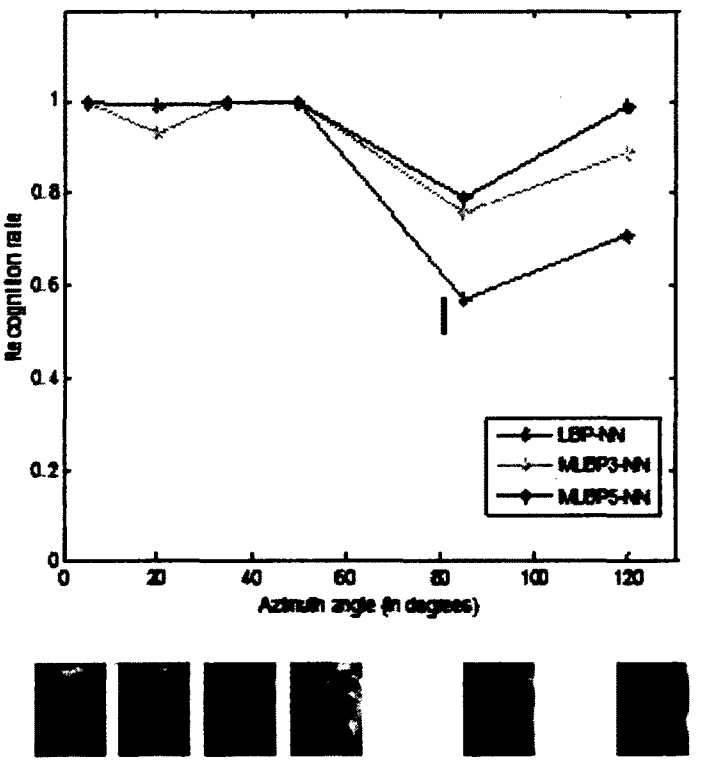


Figure 3.8: Performance comparison of LBP and MLBP with respect to variations of different illumination conditions.

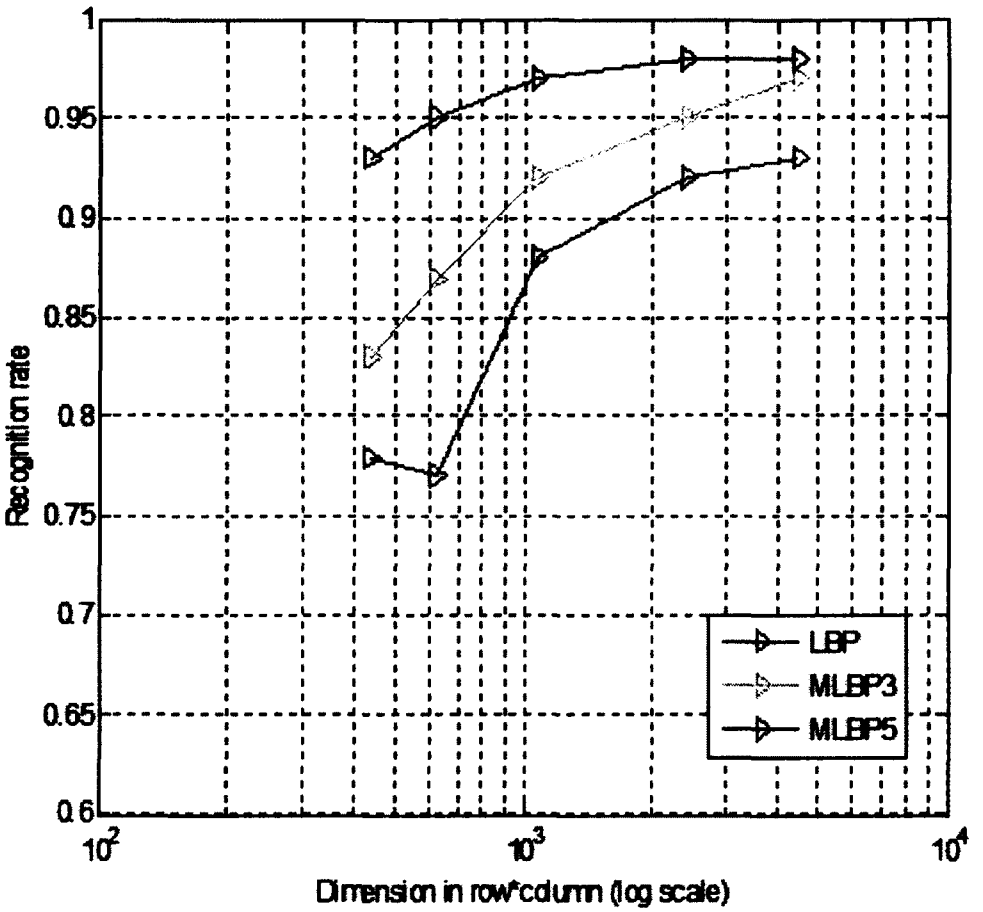


Figure 3.9: Performance comparison of LBP and MLBP with respect to dimensions in row*column vector.

CHAPTER 4 - FACE RECOGNITION USING MARGINALITY PRESERVING EMBEDDING

4.1 Introduction

In many applications of computer vision and machine learning we often need to deal with very high dimensional data but the intrinsic structure of the data may lie in a low dimensional space. Learning such high dimensional data is computationally expensive and not suitable for all practical applications. Moreover, it is also desirable to reduce the dimension for visualization. But this low dimensional data must preserve the underlying structure of high dimensional data in order to be of use. This leads researchers to develop methods of dimensionality reduction that can extract manifold structure of data on which data may reside.

Many techniques (both supervised and unsupervised) for dimensionality reduction have been proposed over the last few decades [50-53]. All these methods have validated one thing in common-recognition rate can significantly be improved at low dimensional subspace. Two of the most primitive techniques for this purpose are Principal Component Analysis (PCA) [42-44] and Linear Discriminant Analysis (LDA) [43-46]. PCA transforms original image space to orthogonal feature space in the sense of mean square error. LDA seeks for linear transformations that minimize within class covariance and maximize between class covariance matrices. Unlike PCA, LDA encodes discriminating features in a linearly separable space that is not necessarily orthogonal (Laplacian). When training data are small PCA can outperform LDA but if the class information is available LDA can be used to find optimal subspace for optimal discrimination [47].

Recently various researches [49-51] focused on face images have shown that data may reside on a nonlinear submanifold. As a result, manifold learning has become popular for face recognition. Some popular nonlinear techniques include Laplacian Eigenmap [47], Locally Linear Embedding (LLE) [49] and Isomap [53]. All these methods showed impressive results on artificial datasets and some real applications. But they are defined only on training data points and it is unclear how the map can be evaluated for new data points. Some linear manifold learning techniques have also been proposed such as Locality Preserving Projection (LPP) [10] and Augmented Relation Embedding (ARE) [13]. LPP uncovers manifold structure by preserving local structure of data while ARE learns manifolds by using users feedback. Yan, *et al.* [14] proposed Marginal Fisher Analysis (MFA) for dimensionality reduction. It redefines LDA in a graph embedding framework by constructing two graphs - one for intraclass compactness and the other for interclass separability. Another graph embedding network called Maximum Margin Projection (MMPP) [15] maximizes the margin between within-class graph and between-class graph. In this paper, a new dimensionality reduction technique called Marginality Preserving Projection (MPP) is proposed. The proposed method considers manifold structure which is modeled by two adjacency graph: one from same class level and the other from local neighborhood. Thus it has more discriminant power and it is defined everywhere.

4.2 Review of Manifold Learning Techniques

4.2.1 Locality Preserving Projections (LPP)

Locality Preserving Projection is a linear approximation of the nonlinear Laplacian Eigenmap [2]. Given a set $\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_m$ in $\mathbf{R}^n$ it finds a transformation matrix $\mathbf{A}$ that map these m points to a set of points $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m$ in R^l $(l << n)$ such that $\mathbf{y}_i = A^T \mathbf{x}_i$. A reasonable criterion for such a map is to minimize the following objective function:

$$\sum_{i,j} (y_i - y_j)^2 W_{i,j} \tag{4.1}$$

where $W_{i,j}$ is a weight matrix constructed in a graph embedding network such that neighboring points are close to each other and the distant points are far away. Let a is a transformation vector. So the objective function can be reduced to

$$\begin{aligned}
&\frac{1}{2}\sum_{i,j} (y_i - y_j)^2 W_{i,j} \\
&= \frac{1}{2}\sum_{i,j} (a^T x_i - a^T x_j)^2 W_{i,j} \\
&= \sum_i a^T x_i D_{ii} x_i^T a \; - \sum_{ij} a^T x_i W_{ij} x_j a \\
&= a^T X (D - W) X^T a \\
&= a^T X L X^T a
\end{aligned} \tag{4.2}$$

The minimization problem is given by adding constraint with the objective function and is given by,

$$\arg \min_{a^T X D X^T a = 1} \quad a^T X L X^T a \tag{4.3}$$

The transformation vector a is given by the eigenvalue solution of the following generalized eigenvalue problem:

$$X L X^T a = \lambda X D X^T a \tag{4.4}$$

4.2.2 Marginal Fisher Analysis (MFA)

Marginal Fisher Analysis overcomes the limitation of LDA that the data of each class are a Gaussian distribution that does not exist in real world applications. Toward this end, two graphs are constructed - one that characterizes intraclass compactness and the other that characterizes interclass separability. By following the graph embedding formulation, intraclass compactness is characterized from the intrinsic graph by the term

$$\tilde{S}_c = \sum_i \sum_{i \in N_{k1}^-(j) \text{ or } j \in N_{k1}^+(i)} \left\| w^T x_i - w^T x_j \right\|^2$$

$$= 2 w^T X (D - W) X^T w \tag{4.5}$$

$$W_{i,j} = \begin{cases} 1, & \text{if } i \in N_{k1}^+(j) \text{ or } j \in N_{k1}^+(i) \\ 0, & \text{else} \end{cases}$$

$$= 2 w^T X \left(D^p - W^p\right) X^T w \tag{4.6}$$

Here $N_{k1}^+(i)$ indicates the index set of the k_1 nearest neighbors of the sample x_i in the same class. Interclass separability is characterized by a penalty graph with the term

$$\tilde{S}_p = \sum_i \sum_{(i,j) \in P_{k2}(c_i) \text{ or } (i,j) \in P_{k2}(c_i)} \left\| w^T x_i - w^T x_j \right\|^2$$

$$= 2 w^T X \left(D^p - W^p\right) X^T w \tag{4.7}$$

$$W_{i,j}^{p} = \begin{cases} 1, & \text{if } (i,j) \in P_{k2}(c_i) \text{ or } (i,j) \in P_{k2}(c_j) \\ 0, & \text{else} \end{cases} \tag{4.8}$$

Here $P_{k2}(c)$ is a set of data pairs that are the k_2 nearest pairs among the set $\{(i,j), i \in \pi_c, j \notin \pi_c\}$. So the marginal Fisher criterion can be given by

$$w = arg \min_{w} \frac{w^T X(D-W)X^T W}{w^T X(D^p - W^p)X^T W} \tag{4.9}$$

4.2.3 Maximum Margin Projection (MPP)

Maximum Margin Projection also constructs two graphs, that is within-class graph G_w and between-class graph G_b. The corresponding weight matrices are given by,

$$W_{b,ij} = \begin{cases} 1, & \text{if } x_i \in N_b(x_j) \text{ or } x_j \in N_b(x_i) \\ 0, & \text{otherwise} \end{cases}$$

$$W_{w,ij} = \begin{cases} \gamma, & \text{if } x_i \text{ and } x_j \text{ share the same level} \\ 1, & \text{if } x_i \text{ or } x_j \text{ is unlabeled} \\ & \text{but } x_i \in N_w(x_j) \text{or } x_j \in N_w(x_i) \\ 0, & \text{otherwise} \end{cases} \tag{4.10}$$

where

$$N_b(x_i) = \{x_i^j \mid l(x_i^j) \neq l(x_i), 1 \leq j \leq k\} \text{ and} \tag{4.11}$$

$$N_w(x_i) = N(x_i) - N_b(x_i) \tag{4.12}$$

Here, $l(x_i)$ is the class level of x_i, and is either relevant or not. For each data point set $N(x_i)$ can be split into two subsets: $N_b(x_i)$ and $N_w(x_i)$. $N_b(x_i)$ contains the neighbors of

different levels and $N_w(x_i)$ contains the rest of the neighbors. So the objective function of Maximum Margin Projection is given by [53]

$$\min \sum_{i,j} (y_i - y_j)^2 W_{w,ij} \tag{4.13}$$

$$\max \sum_{i,j} (y_i - y_j)^2 W_{b,ij} \tag{4.14}$$

The objective function given by (i) and (ii) can be reduced to

$$\min \frac{1}{2}\sum_{i,j} (y_i - y_j)^2 W_{w,ij} = a^T X D_w X^T a - a^T X W_w X^T a \tag{4.15}$$

and

$$\max \frac{1}{2}\sum_{i,j} (y_i - y_j)^2 W_{b,ij} = a^T X L_b X^T a \tag{4.16}$$

where D_w and D_b are diagonal matrices given by

$$D_{w,ii} = \sum_j W_{w,ij} \text{ and } D_{b,ii} = \sum_j W_{b,ij} \tag{4.17}$$

and

$$L_b = D_b - W_b \tag{4.18}$$

is the graph Laplacian matrix of G_b. Now, the objective function in (i) can be written as

$$\min_a 1 - a^T X W_w X^T a \tag{4.19}$$

Equivalently,

$$\max_a a^T X W_w X^T a \tag{4.20}$$

Similarly (ii) can be expressed as

$$\max_a a^T X L_b X^T a \tag{4.21}$$

Finally, the optimization problem can be reduced to finding

$$\arg\max_{a^T X D_w X^T a=1} \; a^T X\left(\alpha L_b + (1-\alpha) W_w\right) X^T a \tag{4.22}$$

where $0<\alpha<1$ is a constant.

4.2.4 Locally Linear Embedding (LLE)

Locally Linear Embedding (LLE) is an unsupervised learning algorithm that computes low-dimensional neighborhood preserving embeddings of high dimensional inputs. LLE maps its inputs into single global co-ordinate systems of lower dimensionality, and its optimizations do not involve local minima. Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$ be the data points sampled from an underlying sub-manifold M embedded in $\Re^n$ and y_i be the one-dimensional map of x_i $i=1,2,....,m$. The basic idea of LLE is to construct a k nearest neighbor graph G with weight matrix W. Reconstructing errors are measured by the following cost function [53]

$$\phi(W)=\sum_{i=1}^{m}\left\| x_i - \sum_{j=1}^{m} W_{ij} x_j \right\|^2 \tag{4.23}$$

which adds up the squared distances between all the data points and their reconstructions. Now consider the mapping of original data points to a line so that each data point on the line can be represented as a linear combination of its neighbors with the coefficients of W_{ij}. Let $y_1, y_2, ..., y_m$ be such a map. A reasonable criterion for choosing a "good" map is to minimize the following cost function:

$$\phi(\mathbf{y})=\sum_{i=1}^{m}\left(y_i - \sum_{j=1}^{m} W_{ij} y_j \right)^2 \tag{4.24}$$

It can be shown that the optimal embedding y is given by the minimum eigenvalue solution to the following eigenvalue problem

$$(I-W)^T (I-W) y = \lambda y \tag{4.25}$$

where I is an $m \times m$ identity matrix.

4.3 Marginality Preserving Embedding (MPE)

The basic idea of Marginality preserving projection (MPE) is to find both geometrical and discriminant features in data manifold. It focuses on Locally Linear Embedding (LLE) in the sense of reconstruction from local neighborhood. But LLE is defined only on the training data and it is not clear how to evaluate new data point. If the number in the training sample is small, reconstruction based on minimizing the error may not be the optimum since the classes with the more frequent examples tend to dominate the prediction of the new vector, as they tend to come up in the k nearest neighbors. This limitation of LLE may be overcome by developing new criteria that minimize the contribution from inter-class samples in reconstruction. Toward this end, a new algorithm MPE is proposed which is based on graph embedding framework that uses both label information and the advantage of LLE to enhance the recognition rate. First, a graph is constructed that uses class information and preserves marginality in reconstruction. Then another graph is constructed that uses both class information and neighborhood information.

Given a set of points $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_m$ in $\Re^n$ the goal is to find a transformation matrix, $\mathbf{A}$ that maps all these points to a set of points $\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_m$ in $\Re^d$ such that $d << n$. Let $W_{i,j}^s$ be a $m \times m$ coefficient matrix of reconstruction from all the members of the same class. The objective function that minimizes the reconstruction error can be defined as

$$\phi(W^s)=\sum_i \left\| x_i - \sum_j W^s_{i,j} x_j \right\|^2 \tag{4.26}$$

which adds up the squared distances between all the data points and their reconstruction. The weights $W^s_{i,j}$ summarize the contribution of the jth data point to the ith reconstruction. The minimization is subjected to two constraints:

$$W^s_{i,j} = 0 \qquad \text{if } x_j \notin C(x_i) \tag{4.27}$$

where $C(x_i)$ is the class of the pattern x_i

and

$$\sum_{x_j \in C(x_i)} W^s_{i,j} = 1 \tag{4.28}$$

Consider the problem of mapping original data point to a line so that each data point on the line can be represented as a linear combination of its class members with coefficients $W^s_{i,j}$. Let $\mathbf{y}=(y_1, y_2, ..., y_m)^T$ be such a map. A reasonable criterion for choosing a "good" map is to minimize the following cost function

$$\phi(\mathbf{y})=\sum_i \left(y_i - \sum_j W^s_{i,j} y_j \right)^2 \tag{4.29}$$

under appropriate constraints. Here the weights $W^s_{i,j}$ are fixed while the coordinate y_i is optimized. Now another cost function is defined, which adds up reconstruction error from neighbors that belong to different class. So, the objective here is to maximize the following cost function:

$$\phi(\mathbf{y})=\sum_i \left(y_i - \sum_j W^d_{i,j} y_j \right)^2 \tag{4.30}$$

with the following constraints

$$W_{i,j}^{d} = 0 \qquad \text{if } x_j \notin N_k(x_i) \text{ and } C(x_i) \neq C(x_j) \tag{4.31}$$

and

$$\sum_{\substack{x_j \in N_k(x_i) \\ x_j \notin C(x_i)}} W_{i,j}^{d} = 1 \tag{4.32}$$

where $N_k(x_i)$ is the k nearest neighbors of x_i and $C(x_k)$ is the class of k th pattern. It can be shown that Equation (4.32) is reduced to

$$\phi(\mathbf{y}) = a^T X M^s X^T a \tag{4.33}$$

where

$$M^s = (I - W^s)^T (I - W^s) \tag{4.34}$$

By combining (3) and (4) objective function finally reduces to solve the following optimization problem

$$\arg\max_{a} \frac{a^T X M^d X^T a}{a^T X M^s X^T a} \tag{4.35}$$

The optimal a's are the eigenvectors corresponding to the following maximum eigenvalue solution:

$$X M^d X^T a = \lambda X M^s X^T a \tag{4.36}$$

It is easy to show that the matrices $X M^d X^T$ and $X M^s X^T$ are symmetric and positive semidefinite.

The algorithmic procedure of Marginality Preserving Projection is described below:

a) **Graph construction:** For the intraclass graph, put an edge between x_i and x_j if they belong to the same class. For the neighborhood graph, put an edge between

x_i and x_j is considered if x_j is among the k nearest neighbors of $x_i$an d they have different class level.

b) **Weight computation:** In this step, weight matrices are computed for each graph. For each pattern a row vector of weights determines the contribution of all samples and is given by,

$$W^s(i,:)=\min\left\|x_i-\sum_j W^s_{i,j}x_j\right\|^2 \tag{4.37}$$

$$W^d(i,:)=\min\left\|x_i-\sum_j W^d_{i,j}x_j\right\|^2 \tag{4.38}$$

c) **Projections:** Compute the eigenvectors and eigenvalues of the generalized eigenvector problem:

$$XM^dX^Ta = \lambda XM^sX^Ta \tag{4.39}$$

where $M^s=\left(I-W^s\right)^T\left(I-W^s\right)$ and $M^d=\left(I-W^d\right)^T\left(I-W^d\right)$.

Let the column vectors $a_0,a_1,...,a_{d-1}$ be the solutions of(4.39) ordered according to their eigenvalues $\lambda_0>\lambda_1>....>\lambda_{d-1}$. Thus the embedding is as follows:

$$\mathbf{x}_i\rightarrow\mathbf{y}_i=A^T\mathbf{x}_i$$

$$A=(a_0,a_1,...,a_{d-1}) \tag{4.40}$$

where $\mathbf{y}_i$ is a d-dimensional vector and A is an $n\times d$ matrix.

4.4 Experiments and Analysis

The ORL (Olivetti Research Laboratory) face database was used in the experiments. The database consists of 10 different images of each of 40 different subjects. The images were taken at different times with varying lighting conditions and facial expressions (open/closed eyes, smiling/not smiling, open/closed mouth) and facial details (glasses/no glasses). All images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some tilting and rotation of face up to 20 degrees). Original images were cropped to locate only the face regions and resized to 32x32 pixels with 256 gray levels/pixel. Thus each face image can be represented by a 1024-dimensional vector. For each subject $n = 2$ and 5 images are used for training and the rest for testing the algorithm. For each n, 50 random selections for training images are used and the mean of the results is estimated. For comparison, the test is run against some popular techniques in face recognition such as Eigenface [1], Fisherface [2], LPP [10] and NPE [9].

Figure 4.1 shows the mapping results for visualizing the effectiveness of MPP. The images are mapped into two dimensional spaces using the first two co-ordinates of NPE and MPP. As can be seen from the figure, the faces of different subjects can be clustered clearly by their MPP mapping structure. This is because MPP not only uses local neighborhood in embedding but it also preserves the uniqueness of each class. Figure 4.2 shows the error rate of different methods against dimensionality. It shows that the recognition rate is affected with number of dimensions in face subspace. From the figure, it can be seen that the proposed MPP attains maximum value at 39 dimensions (for 2 samples) and 30 (for 5 samples). To illustrate the advantage of our method over other

methods the classification accuracy is compared for different numbers of training samples and listed in Table 4.1. As can be seen Fisherface, LPP and NPE have outperformed the baseline and Eigenfaces for both small and large training sizes. But the proposed MPE outperforms all these methods and the improvement is much better in small training sizes.

MPE was also applied to the YaleB database with varying illuminations. The database contains 38 subjects under 64 illumination conditions. It has little variability in pose and expressions but its extreme lighting variations make it a difficult problem in face recognition. From the database 2413 images were selected of 38 individuals. The images are cropped and resized to 36×30 pixels. Two images were used per person for training in the most neutral lighting conditions. From the experiment it was found that without MLBP, MPE performs poorly in this database. The recognition rate is 40.58%. But after applying MLBP recognition rate improves to 76.28%. This shows that this method of face representation and recognition performs well under difficult illumination conditions.

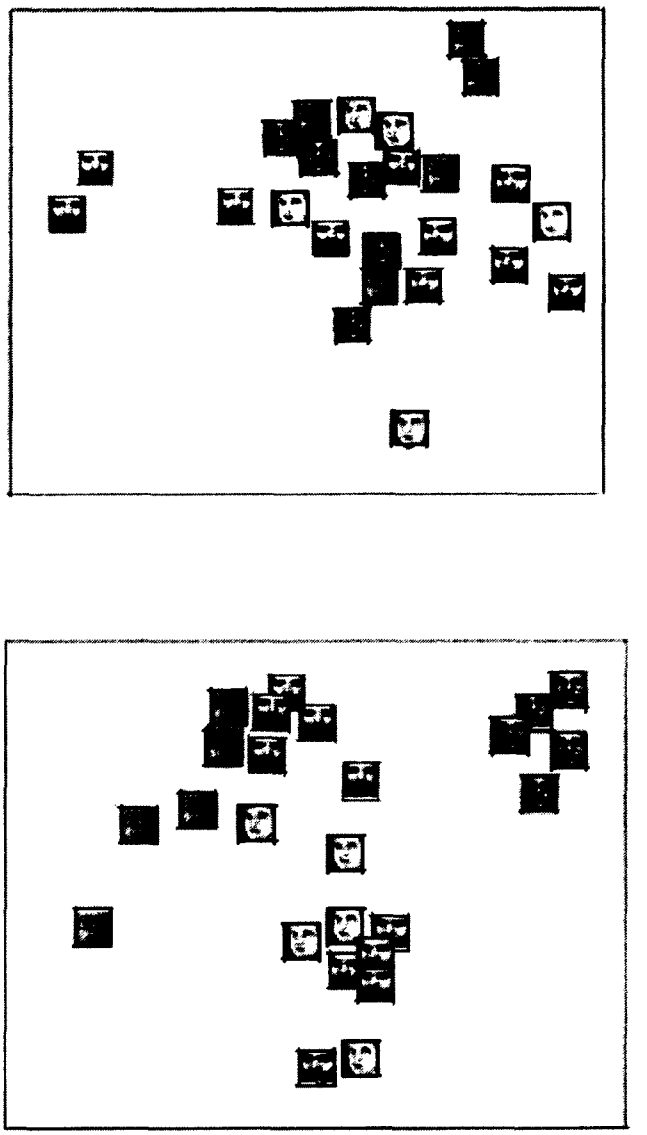


Figure 4.1: A two dimensional representation of the set of face images in test space using NPE (left) and MPE (right). Color frames of each face indicate IDs.

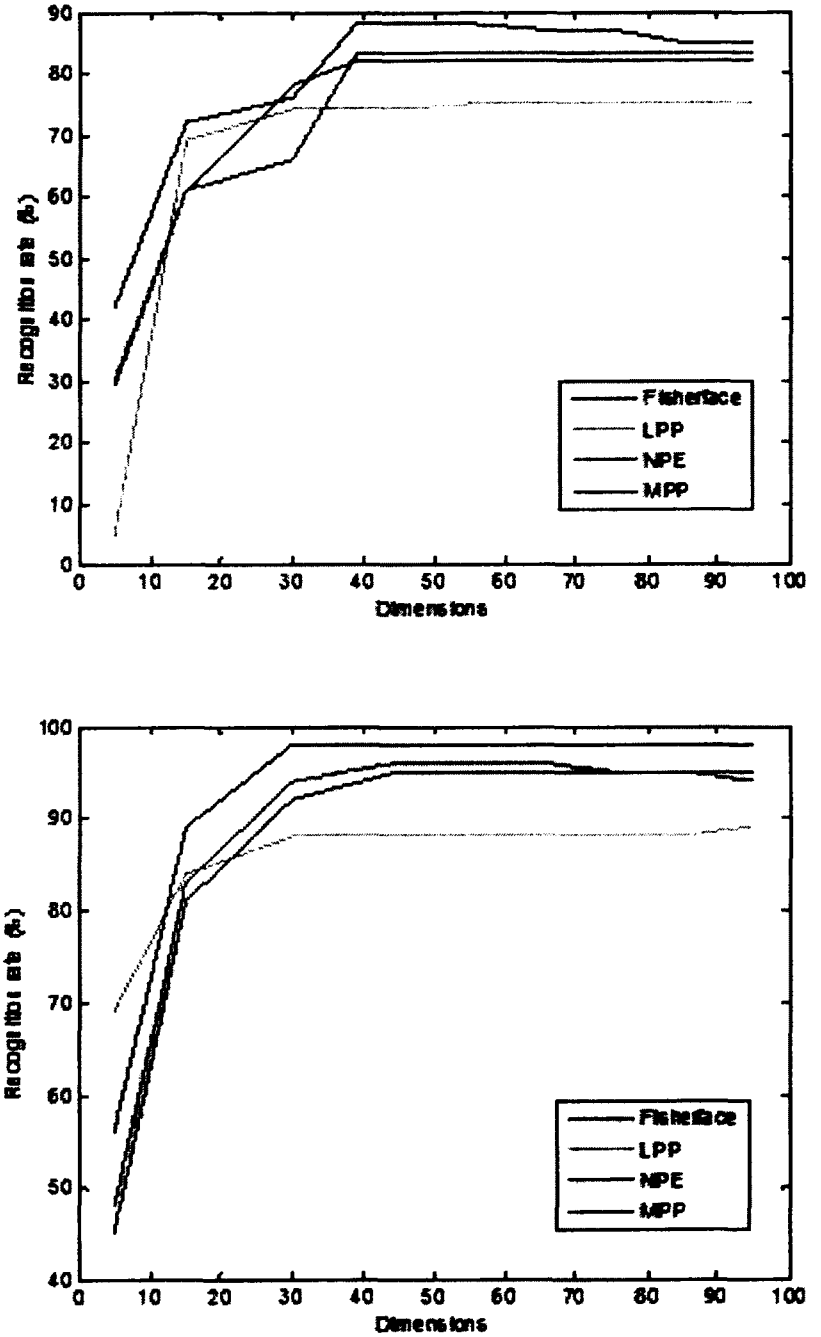


Figure 4.2: Plot of recognition rate with dimensionality on ORL database.

Table 4.1: Face Recognition Results (%) using MPE on ORL database.

Method	2 Samples	5 Samples
Baseline	71	89
Eigenfaces	53	75.14
Fisherfaces	74.6	91.52
LPP	69.22	87.28
NPE	76.04	92.38
MPE	79.51	94.64

4.5 Summary

In this chapter, a new dimensionality reduction technique called Marginality Preserving Embedding (MPE) is proposed. Several other methods also address subspace learning technique for dimensionality reduction both supervised and unsupervised way. Two related algorithms called LLE [8] and NPE [9] also share locality preserving projections. In addition to that, this method also considers similarity and dissimilarity measures by formulating an optimization problem that involves both intraclass and interclass data in the local neighborhood. It is simple and defined everywhere on test data. Performance of this method is demonstrated through several experiments and shows lower error rates in face recognition.

CHAPTER 5 - NEURAL NETWORK BASED FACE RECOGNITION

5.1 Introduction

In this chapter, two novel ideas for face recognition are proposed. First, neural networks are used to train the system and input vectors are created by measuring distance from each input to its class mean. Second, half-face symmetry is used, realizing that face images may contain various expressions open/close eye, open/close mouth, etc. So, the top half and bottom half are separately classified and the results are fused.

Face recognition research has grown rapidly in the past two decades due to security reasons. It is a challenging work because human faces vary in pose, appearance and appear different under varying lighting conditions. Moreover, for real time surveillance applications it is necessary to reduce computation time. A large number of neural network based training systems have been proposed in the literature [68-74]. Early methods of neural network use direct input of training images [68-69]. Then the image feature is first extracted prior to network input. Then comes the feature based input to the neural network. Sudha, et *al.*'s [73] Principal Component Analysis (PCA) neural network first applies PCA to extract the eigenface of the training database. After several iterations, weights are converged to eigenfaces.

5.2 Review of Neural Networks

There are numerous algorithms based on neural networks for face recognition. Some common methods include radial basis network, convolution neural network, etc. All these methods use complex and non-linear learning algorithms. The basic form of multilayer neural network with backpropagation consists of an input layer, one hidden layer, and one output layer. Classification is done through regression by minimizing standard training error between actual output and observed output [76]. A conventional approach of neural network based face recognition uses some kind of feature extraction technique and then feeds it to the input of the classifier.

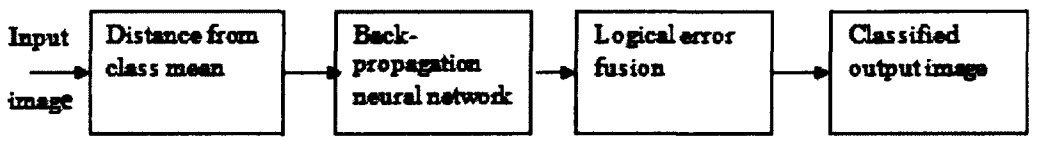


Figure 5.1: Basic block diagram of proposed neural network-based system.

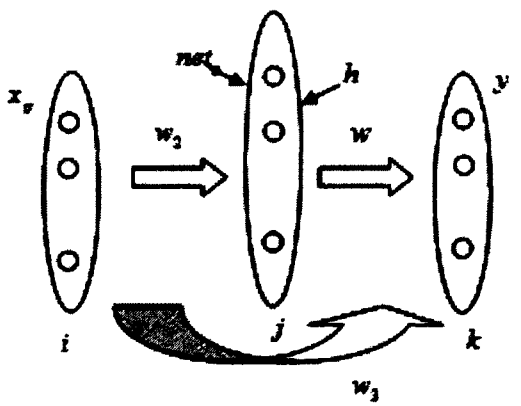


Figure 5.2: Basic block diagram of multi-layer perceptron (MLP) for output weight optimization with back propagation.

Let a set of N_v training patterns (x_p, t_p) where the pth input vector x_p and desired output vector t_p have dimensions N and N_{out} respectively. If jth unit is hidden, unit the net input $net_p(j)$ and output activation function $h(j)$ for pth pattern is given by,

$$net_p(j) = \sum_i w_2(j,i) x_p(i) \tag{5.1}$$

$$h(j) = f\left(net_p(j)\right) \tag{5.2}$$

Here f is a sigmoid activation function given by,

$$f\left(net_p(j)\right) = \frac{1}{1+e^{-net_p(j)}} \tag{5.3}$$

For the kth output unit, output is dependent on input x_p and hidden layer output $h(j)$ as shown in Fig. 5.2 and is given by,

$$y_p(k) = \sum_i w_3(k,i) x_p(i) + \sum_j w(k,j) h(j) \tag{5.4}$$

where $w(k,j)$ is the output weight connecting kth unit to the jth hidden unit and $w_3(k,i)$ is the weight connecting the jth hidden unit to ith input unit. In order to train neural network, the mapping error for the kth output unit is defined as the difference between mapping output and actual output.

$$E(k) = \frac{1}{N_v} \sum_{p=1}^{N_v} \left[t_p(k) - y_p(k)\right]^2 \tag{5.5}$$

For the classification problem, we try to do classification through regression by minimizing the standard training error given in Equation (5.5).

5.3 Proposed Algoirthm

Figure 5.1 shows the block diagram representation of the proposed algorithm. It measures distance of each pattern from all the class means. These vectors are fed to the neural network input. The inputs of training face images are in set $X=[x_1, x_2, ..., x_n$ while the test images are in set $Y=[y_1, y_2, ..., y_n$. Each face image is sized at 32x32. Samples in each class of training set are used to compute class mean. Let there be C classes in the training set and C class means. For each pattern, the Euclidean distance between each pattern and all class means are calculated. So each feature vector is of size C x1. The network is then fed with these feature vectors. Weights are updated for each input according to the learning rule of neural network. A training set is provided as the input until the convergence is attained which is normally done with setting the number of iteration to a large value within which inputs are converged. Each face input is split into two sub-blocks both in horizontal direction - one form (1:16, 1:32) and the other form (17:32, 1:32). The reason behind this is to improve recognition accuracy. Face images usually consist of different expressions such as open/close eye, open/close mouth, wearing sunglass/not wearing sunglass, etc. So, if we take the entire face image for training that could result significant error. With this new technique, the system is first trained and tested using the upper half face and the result of this is stored. Then the lower half face is used to train and test the images. This result is also stored. So, if the image contains sunglass, it can be detected by the lower half face. Similarly, if the image contains open mouth or masks it can be detected by upper half of the face images. The upper and lower part of face error is denoted by e_1 and e_2. Recognition in any part is considered as true

image. So an error is found only when both part fails to detect. In other words, it is the logical *and* between the two errors. Total error is the sum of all errors after logical *and* operation.

$$\text{Error} = \sum_{i} \text{and}(e_1, e_2) \tag{5.6}$$

For the testing phase the given new face image is q is fed as input to the trained network. From this test image all class mean in the training images are subtracted. Then it is projected to trained neural network. The recognition of a new image is done using Euclidian distance between the trained network and the projected test vectors.

5.4 Simulation Results

In this experiment the ORL (Olivetti Research Laboratory) face database was used. The database contains 10 different images of 40 different subjects. The images were taken at different times with varying lighting conditions and facial expressions (open/closed eyes, smiling/not smiling, open/closed mouth) and facial details (glasses/no glasses). All images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some tilting and rotation of face up to 20 degrees). Original images were cropped to locate only the face regions and resized to 32x32 pixels with 256 gray levels/pixel. Thus each face image can be represented by a 1024-dimensional vector. For each subject we use five images for training and the rest for testing the algorithm. We take 50 random splits for training pattern selection. To illustrate the advantage of this method over other methods the classification accuracy is compared for different number of training samples and listed in Table 5.1. As can be seen,

Fisherface, LPP and NPE have outperformed the Eigenfaces but the proposed method outperforms all these methods.

Table 5.1 Face recognition results (%) using neural network on ORL database.

Method	5 Samples
Eigenfaces	75.14
Fisherfaces	91.52
LPP	86.28
NPE	92
Proposed	93

In order to prove that this algorithm is robust under expression variations Yale database was also tested. It contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, with no glasses, normal, right light, sad, sleepy and wink. Here the recognition rate using the MPE algorithm gives 89.69% whereas the neural network based half face similarity matrix gives a 90.91% recognition rate.

CHAPTER 6 - SUMMARY

A novel technique for single image super-resolution has been presented that is especially designed for video surveillance in order to identify distant faces. In other words, its purpose is to enhance face images. The proposed algorithm is a hybrid edge and feature based technique which uses image features in frequency domain. The covariance matrix is estimated component wise - one from learning real parts and the other from imaginary parts and the results are accumulated. The RMSE of the Lena image (Figure x.x.) improved from 14.0422 (for 4X without component learning) to 13.8323 (for 4X with component learning). Unlike other edge based methods, the kernel regression technique was used to interpolate the unknown pixels and the kernel is learned from local features of the image. This enables reduction of noise and artifacts while maintaining the sharpness of the image. Experimental results show that proposed algorithm performs better than other resolution enhancement techniques.

A new technique for image representation and feature extraction was presented, named modified local binary pattern (MLBP) which shows many advantages over original LBP approach. First, it is less sensitive to variations in lighting conditions. Several experiments were conducted by changing lighting conditions and almost in all cases MLBP performed better than LBP in terms of recognition accuracy. Although in some experiments LBP showed better results but the difference is not significant and MLBP is more consistent in all cases. This is because LBP only compares with the center pixel whereas MLBP uses two layer comparisons. It is noted that the recognition accuracy is improved in difficult lighting conditions is based on the magnitude difference

of each pixel from the center pixel. MLBP considers this in every neighborhood of a given pixel in a given patch. This was evident when MLBP5 was used. It performed better than MLBP3. Only two different neighborhood sizes were used in this study, but different neighborrhood sizes can also be used, although there will be a maximum limit on recognition accuracy. In addition, MLBP has better recognition accuracy than LBP at reduced dimensions.

A new dimensionality reduction technique called Marginality Preserving Embedding (MPE) was proposed. Several other papers also addressed subspace learning technique for dimensionality reduction both supervised and unsupervised way. Two related algorithms called LLE [8] and NPE [9] also share locality preserving projections. In addition to that, the method used in this dissertation also considers similarity and dissimilarity measures by formulating an optimization problem that involves both intraclass and interclass data in the local neighborhood. It is simple and defined everywhere on test data. Performance of this method is demonstrated through several experiments and it shows lower error rates in face recognition.

Two novel ideas were also presented for face recognition. First, we introduced distance measure of each pattern from all class means. This is similar to between class separations. Face recognition under varying expressions was also proposed. Facial expressions normally change with the changes in mouth and eye region. A technique which is capable of detecting half face region and then fuses both results which improves recognition accuracy was also proposed.

In this dissertation, face enhancement, representation and two different recognition algorithms was shown. Although low resolution face enhancement does not help in

recognition accuracy, it helps in visualization. Kernel regression based methods were used by covariance estimation. Future work can include dictionary learning to extract high frequency details from example images. This face representation technique, MLBP, performs extremely well in difficult lighting conditions. Other textures feature extraction techniques which can be done in the future and can focus on aging effects because image textures vary if it taken at two different times. Manifold learning techniques can also be improved by manipulating the weight matrix. Finally, face recognition was shown with expression variation although this is challenging work. Only having a half-face matrix is not enough for a recognition task. Future work can improve face recognition when all these constraints are present all together.

REFERENCES

[1] A. Katsaggelos, R. Molina and J. Mateos, *Super Resolution of Images and Video*. Morgan & Claypool, 2007.

[2] M. Elad and A. Feuer, "Restoration of a single super-resolution image from several blurred, noisy and undersampled measured images," *IEEE Transactions on Image Processing*, vol.6, pp. 1646-1658, 1997.

[3] S. P. Kim, N. K. Bose and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1013-1027, 1990.

[4] N. K. Bose, S. P. Kim and H. M. Valenzuela, "Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframes," Proc. *IEEE International Conf. Acoustics, Speech and Signal Processing (ICASSP)*, vol. V, pp. 269-272, 1993.

[5] R. S. Prendergast and T. Q. Nguyen, "A block-based super-resolution for video sequences," *In Proceedings of ICIP*, pp. 1240-1243, 2008.

[6] T. S. Huang and R. Y. Tsay, "Multiple frame image restoration and registration," *in Advances in Computer Vision and Image Processing*, vol. 1, pp. 317-339, 1984.

[7] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.9, pp. 1167-1183, 2002.

[8] S. Carrato, G. Ramponi and S. Marsi, "A simple edge-sensitive image interpolation filter," *In Proceedings of ICIP*, pp. 711-714, 1996.

[9] J. Allebach and P. Wong, "Edge-directed interpolation," *In Proceedings of ICIP*, pp.707-710, 1996.

[10] X. Li and M. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol.10, pp. 1521-1527, 2001.

[11] L. Zhang and X. Wu, "An edge-guided image interpolation algorithm via directional filtering and data fusion," *IEEE Transactions on Image Processing*, vol.15, pp. 2226-2238, 2006.

[12] J. Sun, Z. Xu and H. Shum, "Image super-resolution using gradient profile prior," *In Proceedings CVPR*, pp.1-8, 2008.

[13] S. Dai, M. Han, W. Xu, Y. Wu and Y. Gong, "Soft edge smoothness prior for alpha channel super-resolution," *Proceedings CVPR*, pp. 1-8, 2007.

[14] J. Sun, N. N. Zheng, H. Tao and H. Y. Shum, "Image hallucination with primal sketch prior," *Proceedings CVPR*, pp. 729-736, 2003.

[15] W. Freeman, T. Jones and E. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol.22, pp. 56-65, 2002.

[16] D. Li, S. Simske and R. Mersereau, "Single image super-resolution based on support vector regression," *Proceedings of IJCNN*, pp. 2898-2901, 2007.

[17] H. Chang, D. Yeung and Y. Xiong, "Super-resolution through neighbor embedding," *Proceedings CVPR*, pp. 275-282, 2004.

[18] D. Glasner, S. Bagon and M. Irani, "Super-resolution from a single image," *In Proceedings of ICCV*, pp. 1-8, 2009.

[19] J. Yang, J. Wright, Y. Ma and T. Huang, "Image super-resolution as sparse representation of raw image patches," *Proceedings CVPR*, pp. 1-8, 2008.

[20] Q. Wang, X. Tang and H. Y. Shum, "Patch based blind image super-resolution," *Proceedings of ICCV*, pp. 709-716, 2005.

[21] F. Candocia and J. Principe, "Super-resolution of images based on local correlations," *IEEE Transactions on Neural Networks*, vol. 10, pp. 372-380, 1999.

[22] H. Takeda, S. Farisu and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Transactions on Image Processing*, vol.16, pp. 349-366, 2007.

[23] M. Wand and M. Jones, "Kernel smoothing, ser. monographs on statistics and applied probability. New York: Chapman & Hall, (1995).

[24] K. Weinberger and G. Tesauro, "Metric learning for kernel regression," *International Conference on Artificial Intelligence and Statistics*, pp. 608-615, 2007.

[25] L. Goldmann, U. Monich and T. Sikora, "Components and Their Topology for Robust Face Detection in the Presence of Partial Occlusion," *IEEE Transactions on Information Forensics and Security*, vol. 2, pp. 559-569, 2007.

[26] A. Levin, R. Fergus, F. Durand and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *SIGGRAPH, ACM Transaction on Graphics*, August 2007.

[27] A. Giachetti and N. Asuni, "Fast artifacts-free image interpolation," *Proceedings of BMVC*, 2008.

[28] J. D. Ouwerkerk, "Image Super-resolution Survey," *Image and Vision Computing*, vol 24, pp. 1039-1052, 2006.

[29] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Transactions on Image Processing*, vol. 11, pp. 467-476, April 2002.

[30] Z. Lei, S. Liao, M. Pietikäinen and S. Z. Li, "Face recognition by exploring information jointly in space, scale and orientation," *IEEE Transactions on Image Processing*, vol. 20, pp. 247-256, Jan. 2011.

[31] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary, *Proc. European. Conf. on Computer Vision,* ECCV 2010.

[32] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *Proc. IEEE Computer Soc. Conf. Comput. Vis. Pattern Recognition,* June 1991, pp. 586-591.

[33] J. Yang, D. Zhang, A. F. Frangi and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 131-137, Jan. 2004.

[34] P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs. Fisher faces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Ananlysis and Machine Intelligence*, vol. 19, pp. 711-720, July 1997.

[35] C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Transactions on Neural Networks*, vol. 14, pp. 919-928, July 2003.

[36] P. Comon, "Independent component analysis-a new concept," *Signal Process.* vol. 36, pp. 287-314, 1994.

[37] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, pp. 1635-1650, June. 2010.

[38] T. Ahonen, A Hadid and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 2037-2041, Dec. 2006.

[39] W. Zhang, S. Shan, W. Gao, X. Chen and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," *Proc. IEEE Int. Conf. Computer Vision ICCV,* pp. I: 786-791, 2005.

[40] S. Xie, S. Shan, X. Chen and J. Chen, "Fusing local patterns of Gabor magnitude and phase for face recognition," *IEEE Transactions on Image Processing*, vol. 19, pp. 1349-1361, May. 2010.

[41] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 915-928, June. 2007.

[42] Joliffe, I. *Principal Component Analysis*. Springer-Verlag. 1986

[43] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection". *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol 19 pp. 711-720, 1997.

[44] R. Duda, P. Hart, D. Stork, *Pattern Classification*. Wiley-Interscience, Hoboken, NJ., 2000.

[45] Q. Liu, R. Huang, H. Lu., S. Ma.,Face recognition using kernel based Fisher discriminant analysis. *Fifth International Conf. Automatic Face and Gesture Recognition*, 197-201, 2002.

[46] J. Ye, R. Janardan, C. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems. *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol 26(8), pp. 982-994, 2004.

[47] X,He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, Face recognition using Laplacianface. *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol 27(3), pp.: 328-340, 2005.

[48] A. Martinez, A. Kak, PCA versus LDA. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 23 (2), pp. 228-233, 2001.

[49] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding. *Science,* vol 290, pp. 2323-2326, 2000.

[50] X. He, D. Cai, S. Yan, H. Zhang, Neighborhood preserving embedding. *Proc. 11th Int'l. Conf. Computer Vision,* ICCV 05, 2005.

[51] X. He, P. Niyogi, Locality preserving projections. *Advances in Neural Information Processing Systems,* vol 16, *MIT Press,* 2003.

[52] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering. *Proc. Conf. Advances in Neural Information Processing* Systems, vol 15 pp. #-#, 2001.

[53] J. Tanenbaum, V. Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction. *Science,* vol 290 (22), pp. 2319-2323, 2000.

[51] Y. Lin, T. Liu, H. Chen (2005) Semantic manifold learning for image retrieval. *Proc. 13th Ann. ACM Int'l Conf. Multimedia.*

[52] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 29 (1), pp. 40-51, 2007.

[53] X. He, D. Cai, J. Han, Learning a maximum margin subspace for image retrieval. *IEEE Trans. Knowledge and Data Engineering*, vol 20 (2), pp. 189-201, 2008.

[54] L. Wolf, T. Hassner and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE transactions on Pattern Ananlysis and Machine Intelligence*, vol. 33(10), pp. 1978-1990, 2011.

[55] T. Ojala, M. Pietikäinen and D. Harwood, "A comparative study of texture measures with classification based on feature distribution," *Pattern Recognition*, vol. 29, pp. 51-59, 1996.

[56] T. Ahonen, A. Hadid and M. Pietikäinen, "Face recognition with local binary patterns," *Proc. European. Conf. on Computer Vision,* ECCV, pp. 469-481, 2004.

[57] T. Ojala, M. Pietikäinen and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971-987, July. 2002.

[58] A. Reda and B. Aoued, "Artificial neural network-based face recognition," *Intl symposium on control, communication and signal processing*, pp. 439-442, 2004.

[59] S. Lawrence, C. Giles, A. Tsoi and D. Back, "Face recognition: A convolution neural-network approach," *IEEE Transaction on Neural Networks*, vol. 8, pp. 98-113, 1997.

[60] P. Latha, L. Ganesan and S. Annadurai, "Face recognition using neural network," *Signal Processing: International Journal*, vol. 3, pp. 153-160, YEAR.

[61] Wu. Li, C. Wang, D. Xu and S. Chen, "Illumination invariant face recognition based on neural network ensemble," *IEEE International Conf. on Tools with Artificial Intelligence*, 2004.

[62] J. Zeb, M. Javed and U. Qayyum, "Low resolution single neural network based face recognition," *World Academy of Science, Engineering and Technology*, 2007.

[63] N. Sudha, A. Mohan and P. Meher, "A self-configurable systolic architecture for face recognition system based on principal component neural network," *IEEE Trans. on Circuits, Systems and Video Technology*, vol. 21, pp. 1071-1084, 2011.

[64] K. Diamentaras and S. Kung, "*Principal component neural networks: Theory and applications*, New York: Wiley, 1996.

[65] M. Er, S. Wu, J. Lu and H. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Transaction on Neural Networks*, vol. 13, pp. 697-710, 2002.

[66] S. Xie, S. Shan, X. Chen and J. Chen, "Fusing local patterns of Gabor magnitude and phase for face recognition," *IEEE Trans. On Image Processing*, vol. 19, pp. 1349-1361, 2010.

[67] S. Du and R. Ward, "Improved face representation by nonuniform multilevel selection of Gabor onvolution features," *IEEE Trans. On Systems, Man and Cybernatics Part B*, vol. 39, pp. 1408-1419, 2009.

[68] A. Reda and B. Aoued, "Artificial neural network-based face recognition," *Intl symposium on control, communication and signal processing*, pp. 439-442, 2004.

[69] S. Lawrence, C. Giles, A. Tsoi and D. Back, "Face recognition: A convolution neural-network approach," IEEE Transaction on Neural Networks, vol. 8, pp. 98-113, 1997.

[70] P. Latha, L. Ganesan and S. Annadurai, "Face recognition using neural network," *Signal Processing: International Journal*, vol. 3, pp. 153-160.

[71] Wu. Li, C. Wang, D. Xu and S. Chen, "Illumination invariant face recognition based on neural network ensemble," *IEEE International Conf. on Tools with Artificial Intelligence*, 2004.

[72] J. Zeb, M. Javed and U. Qayyum, "Low resolution single neural network based face recognition," *World Academy of Science, Engineering and Technology*, 2007.

[73] N. Sudha, A. Mohan and P. Meher, "A self-configurable systolic architecture for face recognition system based on principal component neural network," *IEEE Trans. on Circuits, Systems and Video Technology*, vol. 21, pp. 1071-1084, 2011.

[74] K. Diamentaras and S. Kung, "*Principal component neural networks: Theory and applications*, New York: Wiley, 1996.

[75] M. Er, S. Wu, J. Lu and H. Toh, "Face recognition with radial basis function (RBF) neural networks," *IEEE Transaction on Neural Networks*, vol. 13, pp. 697-710, 2002.

[76] M. Manry, M. Dawson and A. Fung, "Fast training of neural networks for remote sensing," Remote Sensing Reviews, vol. 9, pp. 77-96, 1994.

VITA

Mohammad Moinul Islam

CONTACT

Vision Lab
Old Dominion University
E-mail: misla001@odu.edu

EDUCATION

Ph.D. Old Dominion University, (expected, May 2012)
Norfolk, Virginia, USA
Electrical and Computer Engineering, 2012

M.Sc University of South Alabama
Mobile, Alabama, USA
Electrical Engineering, 2007

B.Sc. Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
Electrical and Electronic Engineering, 2002

WORK EXPERIENCE

Research Assistant — 01/08 - Present
Department of Electrical and Computer Engineering,
Old Dominion University, USA
- Resolution enhancement of image and video
- Object detection and recognition

Graduate Teaching and Research Assistant — 08/05 - 12/07
Department of Electrical and Computer Engineering,
University of South Alabama, USA

System Engineer — 06/05 - 07/05
GrameenPhone, Bangladesh

Lecturer — 09/03 - 05/05
American International University Bangladesh, Bangladesh

Lecturer — 11/02 - 08/03
University of Science and Technology Chittagong, Bangladesh

PROFESSIONAL AFFILIATION

- SPIE Student Member
- IEEE Student Member

ACADEMIC HONORS/AWARDS

- Awarded Research Assistantship in the Department of Electrical & Computer Engineering at Old Dominion University, USA since January 2008 to present.
- Awarded Graduate Teaching and Research Assistantship in the Department of Electrical & Computer Engineering at the University of South Alabama, USA from August 2005 to December 2007.
- Received Government Scholarship from the Board of Secondary and Higher Secondary Examination, Bangladesh from 1997 to 2001.